第一章 预备知识

§1 统计判决的基本概念

四十年代末期，A. Wald 提出了一种观点，把统计推断问题看成是人和自然的一种"博奕"．这就是对以后的统计发展起了相当影响的统计判决理论．从这个理论看来，本书的主题——参数估计，不过是一种具有某些特点的统计判决问题．事实证明，这种观点对参数估计的发展，确实起了一定的推动作用．不过，其作用主要在于它开拓了问题的面，而不在于提供了多少解决问题的工具．

为了本书今后的应用，在本节中我们对统计判决理论的若干基本概念作一个简要的介绍．

一、统计判决问题的三要素

为了估计一个未知参数，给出一定的估计量；为了检验某个假设，使用一定的检验函数，等等，它们都是相应的统计问题的解．一般地说，一个统计问题的解是所谓"统计判决函数"．为了定义统计判决函数这一重要概念，需要对构成一个统计判决问题的基本要素给以清楚的描述．这些要素是：样本空间和分布族、行动空间以及损失函数．以下逐点介绍．

1．样本空间和分布族　样本是统计推断所依据的原始资料．从理论上说，样本无非就是具有一定分布(至少是部分未知)的随机变量．在以后，我们常常采用如下的配套记法：若以 X 记样本，则以 x 记 X 的具体观察值(在意义不致含混时，有时样本和样本值的记号不加区分)，而以 $\mathscr{X}$ 记 X 的一切可能值构成的集．为在

数学上处理方便起见，有时 $\mathscr{X}$ 还包含一些 X 的值以外的点．例如，在一定的问题中，X 是只取非负值的随机变量，但为方便计，可以把 $\mathscr{X}$ 取为 R^1(此处及以后，总以 R^n 记 n 维欧氏空间)．

要定义 X 的概率分布，需要给出一个由 $\mathscr{X}$ 的某些子集所构成的 σ-域．我们常把这个 σ-域相应地记为 $\mathscr{B}_x$．那么，可测空间 $(\mathscr{X}, \mathscr{B}_x)$ 就称为(所讨论问题的)样本空间．在 σ-域 $\mathscr{B}_x$ 的意义明确时，我们也称 $\mathscr{X}$ 为样本空间．在绝大多数统计问题中，$\mathscr{X}$ 为有限维欧氏空间 R^n 或其一非空 Borel 子集，在这种情况下，$\mathscr{B}_x$ 总取为 $\mathscr{X}$ 的一切 Borel 子集构成的 σ-域，并称这样的样本空间 $(\mathscr{X}, \mathscr{B}_x)$ 为欧氏的．我们也常说：变量 X 的样本空间为 $(\mathscr{X}, \mathscr{B}_x)$．

如上所述，样本 X 有一定的概率分布，这个分布至少是部分未知的．观察 X 的目的，在于希望从样本观察值 x 得到有关 X 的分布的信息．这样，我们只能假定 X 的分布属于一定的分布族 $\mathscr{P}$．在一定的统计问题里，$\mathscr{P}$ 假定是已知的．自然要提出一个问题：当我们面临一个统计问题时，怎样给出 $\mathscr{P}$？对这个在应用上十分重要但在很大程度上不能由数学来回答的问题，我们只能一般地说：它有时可根据所论问题的专业理论知识来定，有时依赖于以往积累的经验资料，而有时不过是一种纯粹假设，其中考虑在数学上处理简单往往占重要地位．例如，经常采用“正态模型”．在各种问题里，以上三个方面都可能是一种重要的因素．

为了标明分布族 $\mathscr{P}$ 中某一确定的分布，可以对 $\mathscr{P}$ 中的每一分布赋予一个标号，例如 θ．这样，也可以把 $\mathscr{P}$ 写为 $\{P_\theta, \theta \in \Theta\}$，其中 Θ 为一切标号集，当 θ 跑遍 Θ 时，P_θ 跑遍 $\mathscr{P}$．习惯上称 θ 为分布参数．以后，我们总假定在参数值不同时，所对应的分布也不相同．Θ 作为参数 θ 的全部可能值的集，习惯上称为参数空间．通常，Θ 可以理解为这样的一种集，它包含参数 θ 的所有可能值，但不必与后者重合．本书讨论的绝大多数情况是：Θ 为有限维欧氏空间或其一非空 Borel 子集．有时，有必要在 Θ 中引进一个 σ-域 $\mathscr{B}_\Theta$．当 Θ 为欧氏空间或其一 Borel 子集时，$\mathscr{B}_\Theta$ 总取为由 Θ 的一切 Borel 子集构成的 σ-域，即参数(可测)空间 $(\Theta, \mathscr{B}_\Theta)$ 为欧

氏的. 最常见的情况是: 分布族 $\mathscr{P}=\{P_\theta, \theta\in\Theta\}$ 满足如下条件: 存在定义于 $\mathscr{B}_x$ 上的 σ-有限测度 μ, 使对一切的 $\theta\in\Theta$ 有 $\mathscr{P}_\theta\ll\mu$ (这时也记为 $\mathscr{P}\ll\mu$). 于是可把 $\mathscr{P}$ 写成 $\mathscr{P}=\{p_\theta d\mu, \theta\in\Theta\}$ 的形式, 其中 p_θ 为 P_θ 对 μ 的 Radom-Nikodym 导数: $p_\theta(x)=dP_\theta(x)/d\mu(x)$. 以后, 我们常见的情况是, μ 为 Lebesguc 测度, 或为某个可列集上的计数测度, 即对任一集 A, $\mu(A)=A$ 中的点数.

把分布族 $\mathscr{P}=\{P_\theta, \theta\in\Theta\}$ 与样本空间 $(\mathscr{X}, \mathscr{B}_x)$ 配在一起, 构成一个三元组 $(\mathscr{X}, \mathscr{B}_x, \mathscr{P})=(\mathscr{X}, \mathscr{B}_x, P_\theta, \theta\in\Theta)$, 称其为(所讨论的问题的)**概率空间**. 有时也把它称作**样本空间**.

下面考察几个简例:

例 1.1 为估计一物体的重量 b, 对该物体作了 n 次独立称量, 结果记为 $X_1, \cdots, X_n$. 假定每次称量的误差服从正态分布 $N(0, \sigma^2)$, $\sigma^2>0$ 是未知的. 对本问题, 样本为 $X=(X_1, \cdots, X_n)$, 样本空间为 $(R^n, \mathscr{B}^n)$ ($\mathscr{B}^n$ 为 R^n 的一切 Borel 子集构成的 σ-域), 参数空间 Θ 为 $\{\theta:\theta=(b, \sigma), -\infty<b<\infty, \sigma>0\}$, 分布族 $\mathscr{P}=\{P_\theta, \theta\in\Theta\}$, 其中

$$dP_\theta(x)=(\sqrt{2\pi}\,\sigma)^{-n}\exp\left\{-\frac{1}{2\sigma^2}\sum_1^n(x_i-b)^2\right\}dx_1\cdots dx_n.$$

例 1.2 为估计一批产品(有 N 个)的废品率 θ, 从其中随机抽出 n 个, 以 X 记其中的废品个数. 当 $\frac{n}{N}\approx 0$ 或抽样是有放回时, 可以认为 X 服从二项分布 $B(n, \theta)$. 就本例而言, 样本空间自然地取为 $\mathscr{X}=\{0, 1, \cdots, n\}$, 参数空间为 $\Theta=\{\theta:0\leqslant\theta\leqslant1\}$, 分布族可形式地表为

$$dP_\theta(x)=\binom{n}{x}\theta^x(1-\theta)^{n-x}d\mu(x),\ 0\leqslant\theta\leqslant1,$$

其中 μ 为 $\mathscr{X}$ 上的计数测度.

例 1.3 设在例 1.1 中, 假定各次试验的误差均值为 0 且独立同分布, 其它一无所知. 则样本空间仍与例 1.1 一样, 而分布族 $\mathscr{P}$ 则由这样的一些分布 P 构成: $P=F\times F\times\cdots\times F$ (n 重直积), F 是任一有均值的一维分布. 在此, 可以取 F 作为分布 P 的"标号"

即参数(而上面的 P 写为 P_F). 从而有

$$\Theta=\left\{F: F \text{ 为一维分布}, \int_{-\infty}^{\infty}|x|dF(x)<\infty\right\}.$$

本例与前两例不同处在于: 不仅参数空间 Θ 不是欧氏的, 而且分布族 $\mathscr{P}$ 中任一分布 P_F 对"参数值"F 的依赖关系也无法用简单的解析形式表达出来. 一般地, 称前两例(以及与之类似的情况)中的分布族或者说所涉及的统计问题为*参数性的*,而例 1.3 则为*非参数性的*. 但是,这种区别只是一种比较笼统的概念,不可能也无必要划出一条泾渭分明的界限来.

2. *行动空间*(*也称为判决空间*) 对一个统计问题的回答,依问题的性质而异,它可以是一个数(在点估计中)、一个区间(在区间估计中),或者一个决定——接受或拒绝(在假设检验中). 而统计判决问题是一种抽象化了的统计问题,关于它的回答,也就是关于形形色色的具体统计问题的回答的抽象——"判决". 因为一定的判决(如接受或否定某一假设)常导致采取一定的行动, 所以也常把关于统计判决问题的回答称为一个"*行动*". 对一个给定的统计判决问题, 采取何种行动, 这与问题的性质、所获样本以及所采用的"优良性准则"(见后面)有关. 但是,对一定的问题,所可能采取的全部行动,是事先就可能知道的. 这个由全部可能行动构成的集 A 就称为*行动空间*. 有时, 需要将 A 适当扩大一些, 在不少情况下, 有必要引进一个由 A 的某些子集所构成的 σ-域 $\mathscr{B}_A$, 以形成*可测空间* $(A, \mathscr{B}_A)$. 当 A 为有限维欧氏空间或其一 Borel 子集(这是最常见的情况)时, 总把 $\mathscr{B}_A$ 取为由 A 的一切 Borel 子集构成的 σ-域.

例 1.4 在例 1.1 中,若问题是作 b 的点估计, 则行动空间可取为 $A=R^1$; 若要作 b 的区间估计, 则可取 $A=\{a: a=[a_1, a_2], a_1\leqslant a_2\}$. 若问题是同时估计 b 和 σ, 则行动空间可取为 $A=\{a=(a_1, a_2): -\infty<a_1<\infty, a_2>0\}$.

例 1.5 设有 c 个正态总体 $N(a_i, \sigma^2)(i=1, \cdots, c)$. 要求根据对各总体的一些观察结果, 对这 c 个总体的均值 $a_1, a_2, \cdots, a_c$

从小到大排定一个次序. 在此, 每个"行动"都是 1, 2, ⋯, c 的一个置换$(i_1, \cdots, i_c)$, 它表示我们所作的判决是: a_{i_1} 最小, a_{i_2} 次之, ⋯, a_{i_c} 最大. 行动空间 A 由一切这样的置换所构成, 它一共包含 $c!$ 个点.

对假设检验问题而言, 一般可能的行动只有两个(接受, 拒绝), 因此行动空间特别简单.

3. 损失函数 面对一个统计判决问题, 采取一定的行动, 就会招致一定的后果(经济的或其它的). 例如, 在例 1.2 中, 问题是要决定是否接受该批产品, 显然, 当废品率 θ 小时, 对接受这批产品的后果是有利的; 否则, 是不利的.

统计判决理论的一个基本假设, 是认为这种后果最终都可以用数量的形式表达出来, 因此, 若从一定的基点起算, 我们就不妨把后果看成是一种"损失", 它总是非负的, 行动愈接近正确, 损失就愈小; 否则, 就愈大. 这就导致"损失函数"这个重要概念. 损失函数 $L(\theta, a)$ 是一个定义在 $\Theta \times A$ 上的非负函数. 当参数真值为 θ 而采取行动 a 时, 所遭受的损失为 $L(\theta, a)$.

例如, 在例 1.1 中, 若要作 b 的点估计, 则一个常用的损失函数是

$$L(\theta, a)=(b-a)^2 \quad (\theta=(b, \sigma)).$$

这就是著名的平方损失函数, 在本书所讨论的点估计理论中占有特殊重要的地位, 许多重要结果都是在这个损失函数下取得的. 更一般地, 可考虑

$$L(\theta, a)=c(\theta)w(b-a),$$

这里 $c(\theta)>0$, 而 $w(x)$ 为 x 的偶函数, 在 $x \geqslant 0$ 处是非降的. $w(x)$ 为凸函数的情况有相当的重要性, 例如绝对值损失函数, 其中 $w(x)=|x|$.

如果问题是要作 b 的区间估计, 则必须考虑到: 当使用区间 $a=[a_1, a_2]$ 时, 所遭受的损失应从两方面来衡量: 一方面是区间 $[a_1, a_2]$ 是否包含 b; 另一方面是区间的长度. 例如, 可取

$$L(\theta, a)=m[1-I_{[a_1, a_2]}(b)]+(a_2-a_1),$$

其中 $m>0$ 为常数. 或

$$L(\theta,\ a)=|a_1-b|+|a_2-b|$$

等等. 在假设检验问题中, 由于可能行动只有两个, 情况比较简单. 例如, 可采用 0-1 损失函数(决定正确时损失为 0, 否则为 1), 这相当于通常的 Neyman-Pearson 理论.

函数 L 需满足一定的可测性要求. 一般地说, 在行动空间 A 中总是引进了 σ-域 $\mathscr{B}_A$. 这时要求对任意的 $\theta\in\Theta$, $L(\theta,\ a)$ 作为 a 的函数, 为 $\mathscr{B}_A$ 可测. 如果在 Θ 中也引进了 σ-域 $\mathscr{B}_\Theta$, 则 L 要求对 $\mathscr{B}_\Theta\times\mathscr{B}_A$ 可测.

判决函数理论的一个弱点在于: 不仅假定一切行动的后果都可以数量化是不现实的, 即使在这种数量化是合理的情况下, 也往往缺乏足够的根据去选择一个合适的损失函数. 因此, 损失函数的选择不能不带有相当的人为性, 其中数学上的简单可行占有重要地位.

二、判决函数及其风险函数

1. 判决函数 沿用前面的记号. 我们曾提到, 判决函数是统计判决问题的解. 所谓"解", 就是指明所要采取的"行动", 而这个行动应依据所得到的样本 x. 因此, 所谓判决函数, 就是指定义于样本空间 $\mathscr{X}$ 上而取值于 A 内的函数. 若选定了判决函数 δ, 而得到样本 x, 则所采取的行动由 $\delta(x)$ 给出. 对判决函数要求有一定的可测性. 具体地说, 任一判决函数 δ 都必须是由样本空间 $(\mathscr{X},\ \mathscr{B}_x)$ 到行动空间 $(A,\ \mathscr{B}_A)$ 的可测变换.

上面所定义的那种判决函数通常叫做"非随机化的". 所谓"非随机", 是指只要样本 x 一经确定, 则所要采取的行动 $\delta(x)$ 也唯一地规定, 无任何随机性(在得出 x 以后无任何随机性). 更普遍一些, 可以考虑下面情况: 当得出样本 x 时, 并不能唯一地规定所要采取的行动 a, 而只能规定一个在 $\mathscr{B}_A$ 上的概率分布 $\delta(\cdot|x)$, 其意义是: 在得出样本 x 以后, 对任何集 $D\in\mathscr{B}_A$, 采取属于 D 的行动的概率为 $\delta(D|x)$. 我们可以设想, 通过某种随机性的机制来

实现这个概率分布 $\delta(\cdot|x)$. 这样，一经得到样本 x，我们就有了一定的办法最终选出一个行动 a. 因为甚至在得到样本 x 后，a 的决定过程仍有随机性，这样的一种判决函数 $\delta(\cdot|\cdot)$（定义于 $\mathscr{B}_A\times\mathscr{X}$，取值于 $[0,1]$）就称为随机化的. 显然，非随机化判决函数是随机化判决函数的特例. 对随机化判决函数，也有一定的要求：对任何 $D\in\mathscr{B}_A$，$\delta(D|x)$ 作为 x 的函数，为 $\mathscr{B}_x$ 可测的.

随机化判决函数在实用上不方便，且显得不自然. 因此，从应用的观点看，其意义是有限的. 然而，它在理论上确是一个不可缺少的工具. 以后我们会有机会看到这一点.

2. 风险函数　先考虑非随机化的判决函数 δ. 设损失函数为 L. 当得到样本 x 后，采取的行动为 $\delta(x)$. 设真参数值为 θ，则所遭受的损失为 $L(\theta,\delta(x))$. 因为样本 X 有分布 P_θ，平均损失为

$$R(\theta,\delta)=E_\theta[L(\theta,\delta(X))]=\int_{\mathscr{X}}L(\theta,\delta(x))dP_\theta(x),\quad(1.1)$$

$R(\theta,\delta)$ 就是 δ 的风险函数，也就是当采用判决函数 δ，而参数真值为 θ 时，所遭受的平均损失. 应当注意的是：在前面关于 L 和 δ 的可测性的假定下，(1.1)中的积分是有意义的.

在随机化判决函数 δ 的场合，其风险仍定义为平均损失. 这个平均损失要分两步计算：先在固定样本 x 的条件下计算平均损失，结果为 $\int_A L(\theta,a)\delta(da|x)$，然后对 x（在概率分布 P_θ 之下）求平均. 最后得

$$R(\theta,\delta)=\int_x\left[\int_A L(\theta,a)\delta(da|x)\right]dP_\theta(x).\qquad(1.2)$$

同样，在前述关于 L 和 δ 的可测性假定之下，(1.2)中的积分是有意义的.

在现行的判决函数理论中，风险函数是由判决函数所决定的（但应注意，风险函数还与损失函数 L 及分布族 $\{P_\theta,\theta\in\Theta\}$ 有关）最重要的量. 因为现行有关判决函数的优良性准则（见1.3），全都取决于其风险函数.

下面举几个简例说明风险函数的计算：

例 1.6 设在例 1.1 中，问题是要估计方差 σ^2. 这时行动空间自然地取为 $A=\{a: a\geqslant 0\}$，取损失函数为 $L(\theta, a)=(\sigma^2-a^2)^2$ $(\theta=(b, \sigma))$. 对此问题，常用的估计为 $\delta(x)=\frac{1}{n-1}\sum_1^n(x_i-\bar{x})^2$ $\left(\bar{x}=\frac{1}{n}\sum_1^n x_i\right)$. 因为 $(n-1)\delta(x)/\sigma^2\sim\chi^2_{n-1}$，可算出

$$R(\theta, \delta)=E_\theta[\sigma^2-\delta(X)]^2=2\sigma^4/(n-1).$$

若用估计 $\delta_1(x)=\frac{1}{n+1}\sum_1^n(x_i-\bar{x})^2$，则可算出

$$R(\theta, \delta_1)=E_\theta[\sigma^2-\delta_1(X)]^2=2\sigma^4/(n+1).$$

注意，对一切 $\theta=(b, \sigma)$，有 $R(\theta, \delta_1)<R(\theta, \delta)$.

例 1.7 设在例 1.1 中，问题是要作 b 的区间估计. 行动空间如例 1.4 中所述，而损失函数取为

$$L(\theta, a)=1-I_{[a_1, a_2]}(b)+m(a_2-a_1)(\theta=(b, \sigma),\ a=[a_1, a_2]),$$

其中 $m>0$ 为常数. 关于这个问题，最常用的是所谓 t-区间估计

$$\delta(x)=\left[\bar{x}-\frac{s}{\sqrt{n}}t_{n-1, \alpha/2},\ \bar{x}+\frac{s}{\sqrt{n}}t_{n-1, \alpha/2}\right]\triangleq[a_1(x), a_2(x)],$$

其中 $s^2=\frac{1}{n-1}\sum_1^n(x_i-\bar{x})^2$，而 $0<\alpha<1$. 考虑到

$$E_\theta[I_{[a_1(X), a_2(X)]}(b)]=P_{(b, \sigma)}\left\{\left|\frac{\sqrt{n}(\bar{X}-b)}{s}\right|\leqslant t_{n-1, \alpha/2}\right\}=1-\alpha$$

以及
$$E_\theta(s)=\sqrt{\frac{2}{n-1}}\frac{\Gamma(n/2)}{\Gamma((n-1)/2)}\sigma,$$

不难算出

$$R(\theta, \delta)=\frac{2\sqrt{2}m}{\sqrt{n(n-1)}}\frac{\Gamma(n/2)}{\Gamma((n-1)/2)}t_{n-1, \alpha/2}\sigma+\alpha.$$

三、判决函数的优良性准则

在给定了一个统计判决问题时，可以采用的判决函数很多. 因为从原则上说，任一由 $(\mathscr{X}, \mathscr{B}_x)$ 到 $(A, \mathscr{B}_A)$ 的可测变换 δ 都可以用. 例如，为估计例 1.1 中的均值 b，可以用样本均值、样本中位

数以及其他种种估计. 不言而喻，在这众多的可供选择的判决函数中，我们希望选择最好的，或尽可能最好的. 但怎样的判决函数才算是"好"的呢？对这个问题的回答，取决于用什么样的标准来衡量判决函数的优良性. 这就是我们现在要讨论的问题.

1. 一致最优性　既然风险函数可衡量一个判决函数的平均损失，那么自然地，若 δ 与 δ^* 为两个判决函数，有

$$R(\theta, \delta^*) \leqslant R(\theta, \delta) \quad (\text{一切 } \theta \in \Theta) \tag{1.3}$$

成立，则称 δ^* 一致地优于或等同于 δ. 若(1.3)成立，且至少对一个 θ 值有不等号成立，则称 δ^* 一致地优于 δ.

如果一个判决函数 δ^* 一致地优于或等同于其它任一个判决函数 δ，则称 δ^* 为(该判决问题的)一致最优的判决函数. 显然，在以风险函数的大小作为评价判决函数优劣的唯一标准的前提下，一致最优的判决函数是最理想的判决函数. 可惜的是，只有在某些微不足道的情况下，这种一致最优的判决函数才存在. 因此，这个准则直接付诸于实用是没有意义的. 为了克服这个困难，一般使用以下两个方法：其一是事先限制所使用的判决函数必须满足某种合理的要求，以缩小所考查的判决函数的范围，在其中有可能存在一致最优者. 另一种作法是将一致最优准则中的"点点比较"改为某种形式的"整体比较". 这两种情况的实例都会在本书中碰到并起到重要的作用. 下面我们先来介绍几个由一致最优准则派生出来的重要概念.

2. 容许性、完全类、本质完全类　设 δ 为一判决函数. 若存在一个判决函数 δ' 一致地优于 δ，则称 δ 为"不可容许的". 反之，若不存在一致地优于 δ 的判决函数，则称 δ 为"可容许的". 这个概念的直观意义很清楚：若 δ' 一致地优于 δ，那么，我们没有理由使用 δ 而不用 δ'. 换句话说，在使风险尽量小的原则下，不容许使用 δ.

要判定一个统计判决函数是可容许或不可容许的，一般说是一个很困难的问题. 在点估计理论中，对这个问题有不少研究. 本书后面有专章进行讨论.

设 $\mathscr{D}$ 是一判决函数类(例如 $\mathscr{D}$ 是一切判决函数类，或是一切非随机化的判决函数类，一切可表为样本的线性函数的判决函数类，等等)，而 $\mathscr{D}^*\subset\mathscr{D}$ 为 $\mathscr{D}$ 的一个子类．若对任何判决函数 $\delta\in\mathscr{D}-\mathscr{D}^*$，必存在 $\delta^*\in\mathscr{D}^*$，使 δ^* 一致优于 δ，则称 $\mathscr{D}^*$ 构成 $\mathscr{D}$ 的一个完全类．在这种情况下，$\mathscr{D}-\mathscr{D}^*$ 中的判决函数没有被使用的余地．因此，我们可以把考虑的范围由 $\mathscr{D}$ 缩小为 $\mathscr{D}^*$．

更一般地，若对任何 $\delta\in\mathscr{D}$，必存在 $\delta^*\in\mathscr{D}^*$，使 δ^* 一致地优于或等同于 δ，则称 $\mathscr{D}^*$ 构成 $\mathscr{D}$ 的一个本质完全类．在这种情况下，如果只着眼于风险大小，则把所考虑的范围由 $\mathscr{D}$ 缩小为 $\mathscr{D}^*$，不致引起任何损失．所以，从应用的角度看，完全类与本质完全类的效用是一样的．但本质完全类的概念在理论上较为方便，因为证明一个判决函数为本质完全类往往更为容易些．

3．Minimax 准则　设判决函数 δ 的风险函数为 $R(\theta,\ \delta)$，当采用 δ 时，所能遭受的最大风险为

$$M(\delta)=\sup_{\theta\in\Theta}R(\theta,\ \delta).$$

从"防止最坏的情况出现"这个角度考虑，我们可以建立如下的准则：若 $M(\delta')<M(\delta)$，则称 δ' 优于 δ．若 $\mathscr{D}$ 为一判决函数类，$\delta^*\in\mathscr{D}$．若对任何 $\delta\in\mathscr{D}$ 有 $M(\delta)\geqslant M(\delta^*)$，则在这个准则下，没有优于 δ^* 的判决函数，即 δ^* 在这个准则下是最优的，这种 δ^* 称为(所给判决问题的)Minimax 解；在估计问题中则称为 Minimax 估计．

Minimax 准则是从风险函数的整体性质来定义判决函数的优良性的一个例子：对两个判决函数 δ 和 δ'，我们不在每个参数值 θ 处去比较其风险 $R(\theta,\ \delta)$ 和 $R(\theta,\ \delta')$，而只就其一整体性指标 $M(\delta)$ 和 $M(\delta')$进行比较．

4．Bayes 准则　Bayes 学派是在现代统计中很有影响的一个学派．关于这个学派的基本思想和方法，我们将在后面专章中进行讨论．在此只从优良性这个角度作一点很初步的说明．

设有两个判决函数 δ 和 δ'．如前所指出，若对 θ 逐点比较二

者的风险函数 $R(\theta, \delta)$ 和 $R(\theta, \delta')$，往往得不到确定的结论(哪个优哪个劣)．因此，我们可以考虑对 θ 进行某种形式的平均，再比较它们的平均值．

取这种平均最一般的方法，是在 Θ 的一个子 σ-域 $\mathscr{B}_\Theta$ 上给定一个概率测度 $dH(\theta)$，计算

$$B_H(\delta)=\int_\Theta R(\theta, \delta)dH(\theta),\quad B_H(\delta')=\int_\Theta R(\theta, \delta')dH(\theta),$$

它们分别是 $R(\theta, \delta)$ 和 $R(\theta, \delta')$ 以概率测度 dH 为权的加权平均．若 $B_H(\delta')<B_H(\delta)$，则(在 Bayes 准则以及选定 $dH(\theta)$ 之下)称 δ' 优于 δ．若 $\mathscr{D}$ 为一判决函数类，$\delta_H\in\mathscr{D}$，且对任何 $\delta\in\mathscr{D}$ 有 $B_H(\delta)\geqslant B_H(\delta_H)$，则在 Bayes 准则(以及给定 dH)之下，δ_H 为 $\mathscr{D}$ 中的最优者，称它为*在给定"先验分布" dH 之下所给的判决问题的 Bayes 解*．在估计中则称为 *Bayes 估计*．Bayes 准则是从风险函数的整体性质来确定判决函数的优良性的又一个重要例子．

上面我们提到"先验分布"一词，将在第四章中对它作比较充分的说明．这里我们只注意，在上述 Bayes 准则中，有一点不确定的地方，就是在 $\mathscr{B}_\Theta$ 上的分布 dH 应该如何选取．而这个分布是表示各 θ 值在计算 $B_H(\delta)$ 时的重要性大小．因而，从实用的观点来看，它应与各 θ 值在实际问题中出现的频繁程度相一致．例如，某厂产品逐日的废品率一般很低，但偶尔取较大的值．这种情况表明，在取分布 dH 时，更多的概率应集中在 0 点附近．由于这种认识以及由此而确定的分布 dH，往往是基于以往的长期经验，而并非基于手头这一次观察的值，故习惯上称之为*先验分布*．

5．*无偏性和同变性*　有关的理论将在本书的后面详加讨论．在这里我们只针对点估计的情况作一点说明．

无偏点估计的概念，想必读者是了解的：设要估计参数 θ 的某个函数为 $g(\theta)$ (g 已知)而使用估计量 $\delta(x)$．若对分布族 $\mathscr{P}=\{P_\theta, \theta\in\Theta\}$ 中任一分布 P_θ，都有

$$E_\theta[\delta(X)]=\int_{\mathscr{X}}\delta(x)dP_\theta(x)=g(\theta),$$

则称 δ 为 $g(\theta)$ 的一个无偏估计. 如果我们局限于只考虑无偏估计，则所考虑的范围就大大缩小了，而在这大大缩小了的范围内，就有可能存在某种意义下的最优估计. 例如，一致最优无偏估计，它就是在全部无偏估计构成的类中的一致最优估计.

因此，无偏性原则是缩小所考虑的判决函数类的一种方法. 这个缩小是否会有损于问题的实质(更具体地说，就是是否把一些更好的判决函数排除在外了)，取决于在特定的问题中无偏性的原则是否合理或切合实际. 无偏性原则曾经在估计理论中得到广泛的应用，但近年来，由于种种原因(其中包括判决函数理论的影响)，对它的看法已经有所改变了.

同变性[equivalence, 也有称为不变性(invariance)的]的作用与无偏性一样，也是为了缩小所考虑的判决函数的范围. 我们举一个简单的例子来说明这个概念. 设在例 1.1 中要估计 b. 设有样本 $x_1, \cdots, x_n$. 如果我们把度量原点移到 $-c$, 则样本成为 $x_1+c, \cdots, x_n+c$, 而均值 b 转化为 $b+c$. 但问题并无任何变化. 因此，对任一估计量而言，在原来的坐标系下，是用 $\delta(x_1, \cdots, x_n)$ 估计 b, 而在新坐标系下，则是用 $\delta(x_1+c, \cdots, x_n+c)$ 估计 $b+c$. 如果认为上述坐标变换不应影响我们对 b 的估计，则自然地应要求

$$\delta(x_1+c, \cdots, x_n+c)=\delta(x_1, \cdots, x_n)+c \quad (\text{对任何 } c). \tag{1.4}$$

这就是一种同变性. 任一满足条件(1.4)的估计 δ 称为(在所考虑的变换下的)同变估计. 把估计量的范围缩小为一切同变估计，就有可能找到最优者. 例如，一致最优的同变估计.

不难理解，这种缩小所考虑的判决函数的范围的方法是否有损问题的实质，取决于在特定问题中同变性原则是否合理或切合实际，这并不是一个容易回答的问题.

§2 指数型分布族

在本节中，我们要引进和讨论一个重要的分布族——指数型分布族. 这个分布族有以下几个特点：一是它包含了许多常见的重

要分布，象正态分布，二项分布，Poisson 分布，Gamma 分布等等. 二是它有良好的解析性质，因而在数学处理上有很多方便. 第三（这一点特别重要），迄今所获结果表明：许多理论性的统计问题在指数型分布族中有满意的解决，而在其它的分布族则否.

一、定义和例子

设$(\mathscr{X}, \mathscr{B}_x)$为一可测空间，$\mu$为$\mathscr{B}_x$上的非零$\sigma$-有限测度. $T_j(x)\ (j=1, \cdots, k)$为有限的$\mathscr{B}_x$-可测的实函数，Θ为一抽象集合，$Q_j(\theta)\ (j=1, \cdots, k)$为定义于$\Theta$上的有限实函数，满足条件

$$\int_{\mathscr{X}} \exp\left[\sum_1^k Q_j(\theta)T_j(x)\right]d\mu(x)<\infty,\ \theta\in\Theta.$$

记此积分为$\dfrac{1}{c(\theta)}$，则对任何$\theta\in\Theta$，

$$p_\theta(x)d\mu \triangleq c(\theta)\exp\left[\sum_1^k Q_j(\theta)T_j(x)\right]d\mu \tag{2.1}$$

确定一个$\mathscr{B}_x$上的概率分布. 由于表达式(2.1)中出现指数，称分布族$\{p_\theta(x)d\mu,\ \theta\in\Theta\}$为一指数型分布族，简称指数族或指数型族.

在这个定义中没有指明集Θ有何限制. 一般地，它是由一切满足上述条件的θ构成的集. 参见下面的例.

例2.1 设$X_1, \cdots, X$为$iid.$ 随机变量，$X_1\sim N(\theta, 1)$，则$\boldsymbol{X}=(X_1, \cdots, X_n)$的分布为

$$\begin{aligned}p_\theta(x)dx&=(2\pi)^{-n/2}\exp\left[-\frac{1}{2}\sum_1^n(x_i-\theta)^2\right]dx_1\cdots dx_n\\&=(2\pi)^{-n/2}e^{-n\theta^2/2}e^{-n\bar{x}\theta}e^{-\frac{1}{2}\sum_1^n x_i^2}\,dx_1\cdots dx_n\\&\triangleq c(\theta)e^{Q(\theta)T(x)}d\mu(x).\end{aligned}$$

这里 $c(\theta)=(2\pi)^{-n/2}e^{-n\theta^2/2}$，$Q(\theta)=\theta$，$T(x)=n\bar{x}=\sum_1^n x_i$，$d\mu=e^{-\frac{1}{2}\sum_1^n x_i^2}dx_1\cdots dx_n$. 这是指数族，$\Theta$可取为$R^1=(-\infty, \infty)$.

若$X_1, \cdots, X_n$为$iid.$，$X_1\sim N(b, \sigma^2)$，$\theta=(b, \sigma)$，则$\boldsymbol{X}=(X_1, \cdots, X_n)$的分布为

$$p_\theta(x)dx = (2\pi\sigma^2)^{-n/2}\exp\left[-\frac{1}{2\sigma^2}\sum_1^n(x_i-b)^2\right]dx_1\cdots dx_n$$

$$= (2\pi\sigma^2)^{-n/2}\exp\left[-\frac{1}{2\sigma^2}\left(\sum_1^n x_i^2 - 2nb\bar{x} + nb^2\right)\right]dx_1\cdots dx_n$$

$$\triangleq C(\theta)\exp[Q_1(\theta)T_1(x)+Q_2(\theta)T_2(x)]dx_1\cdots dx_n.$$

此处 $$C(\theta) = (2\pi\sigma^2)^{-n/2}\exp(-nb^2/2\sigma^2),$$

$$Q_1(\theta) = -\frac{1}{2\sigma^2},\quad Q_2(\theta) = nb/\sigma^2,$$

$$T_1(x) = \sum_1^n x_i^2,\ T_2(x) = \bar{x},\ d\mu = dx_1\cdots dx_n.$$

这是一个指数族，Θ 可取为 $\{\theta:\theta=(b,\ \sigma),\ -\infty<b<\infty, \sigma>0\}$.

例 2.2 设 $X_1,\ \cdots,\ X_n$ 为 $iid.$，X 服从 Gamma 分布 $G(\alpha,\beta)$，即 X_1 对 L 测度有密度

$$g_{\alpha,\beta}(x_1) = \begin{cases} \dfrac{\alpha^\beta}{\Gamma(\beta)}x_1^{\beta-1}e^{-\alpha x_1}, & x_1>0; \\ 0, & x_1\leqslant 0. \end{cases}$$

记 $\theta=(\alpha,\ \beta)$，$X=(X_1,\ \cdots,\ X_n)$的分布为

$$dP_\theta(x) = \prod_{i=1}^n\left(\frac{\alpha^\beta}{\Gamma(\beta)}x_i^{\beta-1}e^{-\alpha x_i}\right)dx_i$$

$$\triangleq c(\theta)\exp[Q_1(\theta)T_1(x)+Q_2(\theta)T_2(x)]du(x).$$

此处 $c(\theta)=\alpha^{n\beta}/\Gamma^n(\beta)$，$Q_1(\theta)=-\alpha$，$Q_2(\theta)=\beta-1$，而 $T_1(x)=\sum_1^n x_i$，$T_2(x)=\sum_{i=1}^n\log x_i$，$d\mu=dx_1\cdots dx_n$. 样本空间 $\mathscr{X}$ 取为 $\mathscr{X}=\{x=(x_1,\ \cdots,\ x_n):\ x_i>0,\ i=1,\ \cdots,\ n\}$. 至于 Θ，可取为 $\{\theta:\ \theta=(\alpha,\ \beta),\ \alpha>0,\ \beta>0\}$.

例 2.3 设 $X\sim$二项分布 $B(n,\ \theta)$. 它可写为

$$dP_\theta(x) = \binom{n}{x}\theta^x(1-\theta)^{n-x}d\mu$$

$$\triangleq c(\theta)e^{Q(\theta)T(x)}d\mu,$$

其中 $c(\theta)=(1-\theta)^n$，$Q(\theta)=\log\dfrac{\theta}{1-\theta}$，$T(x)=x$. 测度 μ 定义在样本空间 $\mathscr{X}=\{0,\ 1,\ \cdots,\ n\}$ 上，$\mu(\{x\})=\binom{n}{x}$，$x=0,\ 1,\ \cdots,\ n$.

本例中 Θ 最大可取为 $\{\theta:0<\theta<1\}$. 而在二项分布族中, $\theta=0$ 和 $\theta=1$ 两个值也是容许的. 这样, 可看到如下情况: 为了将一个分布族看成指数族, 有必要牺牲某些参数值而将分布族缩小一点. 这个情况在 Poisson 分布族中也出现: X 的分布为

$$dP_\theta(x)=\frac{1}{x!}e^{-\theta}\theta^x d\mu \triangleq c(\theta)e^{Q(\theta)T(x)}d\mu.$$

其中 $c(\theta)=e^{-\theta}$, $Q(\theta)=\log\theta$, $T(x)=x$, 测度 μ 定义在样本空间 $\mathscr{X}=\{0, 1, 2, \cdots\}$ 上, 为

$$\mu(\{x\})=\frac{1}{x!},\ x=0, 1, 2, \cdots.$$

Θ 最大可取为 $\{\theta:\theta>0\}$, 而在原来的 Poisson 分布族中, $\theta=0$ 也是容许的.

例 2.4 设 n 维随机向量 $X=(X_1, \cdots, X_n)$ 服从 n 维正态分布 $N(d, A^{-1})$. 我们记 $\theta=(d, A)$, 此处 d 为 n 维实向量, 而 A 为 n 阶正定方阵, X 的分布为

$$\begin{aligned}dP_\theta(x)&=(2\pi)^{-n/2}(\det A)^{\frac{1}{2}}\exp\left[-\frac{1}{2}(x-d)'A(x-d)\right]dx_1\cdots dx_n\\&=(2\pi)^{-n/2}(\det A)^{1/2}\exp\left\{-\frac{1}{2}d'Ad\right\}\\&\quad\times\exp\left\{d'Ax-\frac{1}{2}x'Ax\right\}dx_1\cdots dx_n\\&=c(\theta)\exp\left(\sum_1^q Q_i(\theta)T_i(x)\right)d\mu.\end{aligned}$$

此处 $c(\theta)=(2\pi)^{-n/2}(\det A)^{1/2}\exp\left(-\frac{1}{2}d'Ad\right)$, $q=n+\frac{n(n+1)}{2}$, $(T_1(x), \cdots, T_q(x))=(x_1, \cdots, x_n, x_1^2, \cdots, x_n^2, x_1x_2, \cdots, x_{n-1}x_n)$. 相应地, $Q_1(\theta), \cdots, Q_q(\theta)$ 也不难写出, $d\mu=dx_1\cdots dx_n$. 因此这是一个指数族.

指数族的一个性质是族中的分布有共同的负荷集. 这个集显然就是测度 μ 的负荷集. 因此, 任何一个不具备这个性质的分布族, 例如均匀分布族 $\{R(0, \theta):\theta>0\}$, 都不是指数族.

一般形式的指数族可以转化为一种较简单的标准形式. 为

此，只需改变标号：令 $\theta'_i=Q_i(\theta), i=1, \cdots, k$; $\theta'=(\theta'_1, \cdots, \theta'_k)$, (2.1)就变为

$$p'_{\theta'}(x)d\mu \triangleq c'(\theta')\exp\left[\sum_1^k \theta'_j T_j(x)\right]d\mu.$$

将 p', θ', c' 等上的"'"除去，即得指数族的标准形式

$$p_\theta(x)d\mu=c(\theta)\exp\left[\sum_1^k \theta_j T_j(x)\right]d\mu. \tag{2.2}$$

在这个标准形式下，参数空间一般都取为

$$\Theta=\left\{\theta:\theta=(\theta_1, \cdots, \theta_k), \int_{\mathscr{X}}\exp\left[\sum_1^k \theta_j T_j(x)\right]d\mu<\infty\right\},$$

它被称为自然参数空间.

例 2.5 取样本空间 $\mathscr{X}=R^1$. 又设 $a>0$ 和 b 为已知常数. 考虑分布族

$$e^{b\theta^3/a}\frac{a^2-\theta^6}{2a}\exp[-|ax+b|+\theta^3 x]dx,$$

这可变换为如下标准形式：

$$p_\theta(x)d\mu=c(\theta)e^{\theta x}d\mu(x),$$

其中 $c(\theta)=e^{b\theta/a}\dfrac{a^2-\theta^2}{2a}$, $d\mu(x)=e^{-|ax+b|}dx$, 自然参数空间为 $\Theta=\{\theta:|\theta|<a\}$.

更进一步，我们可以通过变换

$$x'_i=T_i(x) \quad (i=1, \cdots, k), \tag{2.3}$$

把(2.2)化为更简单的形式. 与对 θ 的变换不同之处在于：这个变换要改变原来的样本空间. 这样，得到最简单的指数族形式：

$$p_\theta(x)d\mu=c(\theta)\exp\left[\sum_1^k \theta_i x_i\right]d\mu(x). \tag{2.4}$$

这个形式已在理论研究中被广泛的应用.

二、指数型分布族的基本性质

定理 2.1 指数族(2.4)的自然参数空间为凸集.

证 设 θ、$\theta^*\in\Theta$, $0<p<1$, $q=1-p$. 由 Hölder 不等式，有

$$\int_{\mathscr{X}}\exp\left(\sum_1^k(p\theta_j+q\theta_j^*)x_j\right)d\mu$$
$$=\int_{\mathscr{X}}\left[\exp\left(\sum_1^k\theta_jx_j\right)\right]^p\left[\exp\left(\sum_1^k\theta_j^*x_j\right)\right]^q d\mu$$
$$\leqslant\left[\int_{\mathscr{X}}\exp\left(\sum_1^k\theta_jx_j\right)d\mu\right]^p\left[\int_{\mathscr{X}}\exp\left(\sum_1^k\theta_j^*x_j\right)\right]^q<\infty,$$

其中 $\theta=(\theta_1, \cdots, \theta_k)$，$\theta^*=(\theta_1^*, \cdots, \theta_k^*)$. 这就证明了所要的结果.

特别，在 θ 为一维时，由本定理知，自然参数空间必为一个区间(有限或无限，开、闭，或半开半闭均可)；在高维时，自然参数空间的边界可呈现复杂的情况. 显然，本定理的结论及证明对指数族(2.2)也成立.

考虑指数族(2.4). 自然参数空间 Θ 是 R^k 之一个子集. 由 Θ 的一切内点所构成的集称为 Θ 的内部，用符号 Θ^0 记之.

定理 2.2 设 $\varphi(x)$ 为 $\mathscr{B}_x$-可测函数，满足条件：对一切的 $\theta=(\theta_1, \cdots, \theta_k)\in\Theta^0$，有

$$\int_{\mathscr{X}}|\varphi(x)|\exp\left(\sum_1^k\theta_jx_j\right)d\mu<\infty.$$

将 θ_j 视为复变数，即 $\theta_j=\xi_j+i\eta_j$ $(j=1, \cdots, k)$，则有

(i) 积分

$$f(\theta)=\int_{\mathscr{X}}\varphi(x)\exp\left(\sum_1^k\theta_jx_j\right)d\mu \tag{2.5}$$

在区域

$$\Theta^*=\{\theta=(\theta_1, \cdots, \theta_k):\theta_j=\xi_j+i\eta_j,$$
$$j=1, \cdots, k, (\xi_1, \cdots, \xi_k)\in\Theta^0\}$$

内每点有意义，且为 $\theta_1, \cdots, \theta_k$ 的解析函数.

(ii) 在区域 Θ^* 内，f 对 $\theta_1, \cdots, \theta_k$ 的任意阶偏导数可通过(2.5)式在积分号下求导得出.

证 (i)的前半由

$$\left|\varphi(x)\exp\left(\sum_1^k\theta_jx_j\right)\right|=|\varphi(x)|\exp\left(\sum_1^k\xi_jx_j\right)$$

及 $(\xi_1, \cdots, \xi_k)\in\Theta^0$ 得出. (i)的后半可与(ii)结合起来证明. 为

确定计，考虑 $f(\theta)$ 对 θ_1 的导数，在 Θ^* 中取定 $\theta^0=(\theta_1^0,\ \cdots,\ \theta_k^0)=(\xi_1^0+i\eta_1^0,\ \cdots,\ \xi_k^0+i\eta_k^0)$，记 $\theta'=(\theta_1',\ \theta_2^0,\ \cdots,\ \theta_k^0)$，此处 $\theta_1'=\xi_1'+i\eta_1'$。由于 $(\xi_1^0,\ \cdots,\ \xi_k^0)\in\Theta^0$，存在充分小的 $\delta>0$，使

$$|\xi_1'-\xi_1^0|<\delta\Rightarrow(\xi_1',\ \xi_2^0,\ \cdots,\ \xi_k^0)\in\Theta^0.$$

因为当 $|\theta_1-\theta_1^0|<\delta$ 时，有 $|\xi_1'-\xi_1^0|<\delta$，故这时有 $\theta'\in\Theta^*$。记 $\psi(\theta)=\varphi(x)\exp\left(\sum_1^k\theta_jx_j\right)$，有

$$\begin{aligned}
&|(\psi(\theta')-\psi(\theta^0))/(\theta_1'-\theta_1^0)|\\
&=\left|\varphi(x)\exp\left(\sum_1^k\theta_j^0x_j\right)\right||\exp((\theta_1'-\theta_1^0)x_1)-1|/|\theta_1'-\theta_1^0|\\
&=\left|\varphi(x)\exp\left(\sum_1^k\xi_j^0x_j\right)\right||\exp((\theta_1'-\theta_1^0)x_1)-1|/|\theta_1'-\theta_1^0|\\
&\leqslant|\varphi(x)|\exp\left(\sum_1^k\xi_j^0x_j\right)\exp(|x_1|\delta)/\delta\quad(\text{当 } |\theta_1'-\theta_1^0|<\delta).
\end{aligned}\tag{2.6}$$

最后一步用了如下的不等式：当 $|z|<\delta$ 时，有

$$|(e^{az}-1)/z|\leqslant e^{\delta|a|}/\delta.$$

这不难通过将 e^{az} 展开为级数以验证之。又因

$$e^{|x_1|\delta}\leqslant e^{-x_1\delta}+e^{x_1\delta},$$

由(2.6)得

$$\begin{aligned}
&|(\psi(\theta')-\psi(\theta^0))/(\theta_1'-\theta_1^0)|\\
&\leqslant|\varphi(x)|\left\{\exp\left[(\xi_1^0-\delta)x_1+\sum_2^k\xi_j^0x_j\right]\right.\\
&\quad\left.+\exp\left[(\xi_1^0+\delta)x_1+\sum_2^k\xi_j^0x_j\right]\right\}.
\end{aligned}$$

因为 $(\xi_1^0\pm\delta,\ \xi_2^0,\ \cdots,\ \xi_k^0)$ 都属于 Θ^0，由定理假定知，上式右边对 μ 可积，因而令 $\delta\to0$，并用控制收敛定理，即得 $\partial f(\theta)/\partial\theta_1|_{\theta=\theta^0}$ 存在且等于

$$\int_{\mathscr{X}}\varphi(x)x_1\exp\left(\sum_1^k\theta_j^0x_j\right)d\mu.$$

使用类似的方法，并利用归纳法，即可证明 $f(\theta)$ 对 $\theta_1,\ \cdots,\ \theta_k$ 的各阶偏导数存在且可在积分号下求导。定理证毕。

显然，本定理的结论及证法，对指数族(2.2)也成立．由本定理可得到下面几个有用的推论．

系 2.1 对指数族(2.2)，在区域 Θ^* 内，$c(\theta)$为 $\theta_1, \cdots, \theta_k$ 的解析函数．

证明 由定理 2.2 可直接得出．

系 2.2 关于指数族(2.2)，对任何 $\theta\in\Theta^0$，有

$$E_\theta[T_j(X)] = -\partial\log c(\theta)/\partial\theta_j \quad (j=1, \cdots, k); \tag{2.7}$$

$$\operatorname{cov}_\theta(T_i(X), T_j(X)) = -\partial^2\log c(\theta)/\partial\theta_i\partial\theta_j \quad (i、j=1, \cdots, k). \tag{2.8}$$

证 用定理 2.2，在等式

$$1=c(\theta)\int_{\mathscr{X}}\exp\left(\sum_1^k\theta_tT_t(x)\right)d\mu$$

两边对 θ_j 求导，得

$$\begin{aligned}0&=\frac{\partial c(\theta)}{\partial\theta_j}\int_{\mathscr{X}}\exp\left(\sum_1^k\theta_tT_t(x)\right)d\mu\\&\quad+c(\theta)\int_{\mathscr{X}}T_j(x)\exp\left(\sum_1^k\theta_tT_t(x)\right)d\mu\\&=\frac{\partial c(\theta)}{\partial\theta_j}\cdot\frac{1}{c(\theta)}+E_\theta[T_j(X)]\\&=\frac{\partial\log c(\theta)}{\partial\theta_j}+E_\theta[T_j(X)].\end{aligned}$$

这就证明了(2.7)式．由(2.7)得

$$\begin{aligned}&\frac{\partial}{\partial\theta_i}\left(\frac{\partial\log c(\theta)}{\partial\theta_j}\right)\\&=\frac{\partial}{\partial\theta_i}\left(-c(\theta)\int_{\mathscr{X}}T_j(x)\exp\left(\sum_1^k\theta_tT_t(x)\right)d\mu\right)\\&=-\frac{\partial c(\theta)}{\partial\theta_i}\cdot\frac{E_\theta[T_j(X)]}{c(\theta)}\\&\quad-c(\theta)\int_{\mathscr{X}}T_i(x)T_j(x)\exp\left(\sum_1^k\theta_tT_t(x)\right)d\mu\\&=E_\theta[T_i(X)]E_\theta[T_j(X)]-E_\theta[T_i(X)T_j(X)]\\&=-\operatorname{cov}_\theta(T_i(X), T_j(X)).\end{aligned}$$

这就证明了(2.8)式．

§3 凸集与凸函数

在参数估计理论中，凸损失函数有较大的重要性．因此，在一些理论问题中，会碰到与凸性有关的若干事实．这个内容本不属于本书的范围，但是，我们考虑到有关的材料往往比较分散而不便查找，且今后用到的个别特殊结果在一般书本中大多未加讨论．为了方便读者，我们在本节中把关于凸性的有关事实作一简略的介绍．

一、R^n 中的凸集

设 $C\subset R^n$．若对任何 x、$y\in C$ 及 $0\leqslant p\leqslant 1$，有 $px+(1-p)y\in C$，则称 C 为凸集．

以下对 R^n 中任意两点 $x=(x_1, \cdots, x_n)$ 和 $y=(y_1, \cdots, y_n)$ 以 (x, y) 记它们的内积 $\sum_1^n x_i y_i$．(x, x) 有时记为 $\|x\|^2$．对任何 $x\in R^n$ 及 $\varepsilon>0$，以 $V_\varepsilon(x)$ 记集 $\{y: \|y-x\|<\varepsilon\}$．

引理 3.1 设 C 为凸集，$y\in R^n$，且存在 $\varepsilon>0$ 使 $C\cap V_\varepsilon(y)=\emptyset$．则存在 $a\in R^n$，$\|a\|=1$，且

$$\inf_{x\in C}(a, x)>(a, y). \tag{3.1}$$

证 用 $\bar{C}$ 记 C 的闭包．因为 $\varphi(x)=\|x-y\|^2$ 在 $\bar{C}$ 上连续，故存在 $b\in\bar{C}$，使 $\varphi(b)=\inf\limits_{x\in\bar{C}}\varphi(x)$．由于 C 为凸集，易见 $\bar{C}$ 也是凸集，故若 $x\in\bar{C}$ 而 $0\leqslant p\leqslant 1$，则有 $(1-p)x+pb\in\bar{C}$，因而

$$\|(1-p)x+pb-y\|^2\geqslant\|b-y\|^2.$$

将此式展开并经过化简，即得

$$(1-p)\|x-b\|^2+2(x-b, b-y)\geqslant 0.$$

再令 $p=1$ 即得 $(x-b, b-y)\geqslant 0$．又由假定，$b\neq y$，故 $(b-y, b-y)>0$．因而

$$(x, b-y)\geqslant(b, b-y)>(y, b-y).$$

记 $a=(b-y)/\|b-y\|$，则 $\|a\|=1$，且对一切的 $x\in\bar{C}$，有 $(a, x)>$

(a, y). 因为(a, x)为x在$\bar{C}$上的连续函数，故得(3.1). 引理证毕.

由这个引理容易证明关于凸集的一个重要结果. 先回忆几个名词：若$C\subset R^n$而$x\in R^n$，则当存在$\varepsilon>0$使$V_\varepsilon(x)\subset C$时，称$x$为$C$的内点；若对任给的$\varepsilon>0$，$V_\varepsilon(x)\cap C$与$V_\varepsilon(x)\cap(R^n-C)$均不空，则称$x$为$C$的边界点. 这时必有$x\in\bar{C}$，但未必有$x\in C$. 又设$C$为凸集，且对任何$x$、$y\in\bar{C}$及$0<p<1$，必有$px+(1-p)y$为$C$的内点，则称$C$为严凸集.

定理 3.1 设$C\subset R^n$为凸集，y为其边界点，则存在$a\in R^n$，$\|a\|=1$，使对任意的$x\in\bar{C}$，有

$$(a,\ x-y)\geqslant 0; \tag{3.2}$$

且当C为严凸时，(3.2)式中的等号当且仅当$x=y$时才成立.

这个定理的几何意义是：存在一个过y点的超平面(其方程为$(a,\ x-y)=0$)，使$\bar{C}$全落在此超平面的一侧. 常称此超平面为C过y点的支撑平面.

证 取一串$y_n\bar{\in}\bar{C}$，但$y_n\to y(n\to\infty)$. 由引理3.1，对每个n存在a_n，使$\|a_n\|=1$，且对任意的$x\in\bar{C}$，有

$$(a_n,\ x)>(a_n,\ y_n).$$

因为$\{a_n\}$有界，可取出其一收敛子列. 不失普遍性，设$\lim\limits_{n\to\infty}a_n=a$，有$\|a\|=1$. 对任何$x\in\bar{C}$，有

$$(a,\ x)=\lim_{n\to\infty}(a_n,\ x)\geqslant\lim_{n\to\infty}(a_n,\ y_n)=(a,\ y).$$

即证明了(3.2). 由(3.2)可知，对C的内点x，有$(a,\ x-y)>0$. 事实上，若$(a,\ x-y)=0$，因为x为内点，当$\varepsilon>0$充分小时，$x_\varepsilon=x-\varepsilon a\in C$，但

$$(a,\ x_\varepsilon)=(a,\ x)-\varepsilon(a,\ a)=(a,\ y)-\varepsilon<(a,\ y).$$

这与(3.2)矛盾.

现设C为严凸. 若存在$x\in\bar{C}$，$x\neq y$，使$(a,\ x-y)=0$，则对于任何$p\in(0,\ 1)$，令$z=(1-p)x+py$，则有

$$(a,\ z-y)=(1-p)(a,\ x-y)=0.$$

于是由上面证明的事实，知 z 不能是 C 的内点，这与 C 的严凸性相矛盾．定理证毕．

二、凸 函 数

设 $C\subset R^n$ 是凸集，$f(x)$ 为定义于其上的有限实函数．若对 C 中任意两点 x、y，恒有

$$f\left(\frac{x+y}{2}\right)\leqslant\frac{1}{2}[f(x)+f(y)],\tag{3.3}$$

则称 f 为(定义于 C 上的) 凸函数．又若当 $x\neq y$ 时 (3.3) 中必有不等号成立，称 f 为严凸函数．

从定义易知：若 f、g 都是凸函数，则 $f+g$ 也是．又若 f、g 中至少有一个为严凸，则 $f+g$ 也为严凸．若 f 为(严)凸，而 $a>0$ 为常数，则 af 也为(严)凸．若对任何 $x\in C$ 有 $\lim\limits_{n\to\infty}f_n(x)=f(x)$ 且当 n 充分大时 f_n 为 C 上的凸函数，则 f 也是 C 上的凸函数．

引理3.2 设 f 为定义在凸集 $C\subset R^n$ 上的有限实函数．则 f 为(严)凸的充要条件是：对于任意的 x_1、$x_2\in C$，$x_1\neq x_2$，$\varphi(t)=f((1-t)x_1+tx_2)$ 是变量 t 在区间 $[0,1]$ 上的(严)凸函数．

证 由 C 为凸集知 $\varphi(t)$ 在 $0\leqslant t\leqslant 1$ 上有定义．若 f 为凸，则对 t、$t'\in[0,1]$，有

$$\begin{aligned}\varphi\left(\frac{1}{2}(t+t')\right)&=f\left(\frac{1}{2}(((1-t)x_1+tx_2)+((1-t')x_1+t'x_2))\right)\\&\leqslant\frac{1}{2}f((1-t)x_1+tx_2)+\frac{1}{2}f((1-t')x_1+t'x_2)\\&=\frac{1}{2}(\varphi(t)+\varphi(t')).\end{aligned}$$

反过来，若 $\varphi(t)$ 在 $0\leqslant t\leqslant 1$ 为凸，则对任何 x_1、$x_2\in C$，有

$$f\left(\frac{1}{2}(x_1+x_2)\right)=\varphi\left(\frac{1}{2}\right)\leqslant\frac{1}{2}(\varphi(0)+\varphi(1))=\frac{1}{2}(f(x_1)+f(x_2)).$$

这就证明了引理中凸的部分，严凸部分的证明是一样的．

引理3.3 设 f 为定义在凸集 C 上的凸函数，则对任意的自然数 n 以及 C 中任意 n 个点 $x_1,\cdots,x_n$，有

$$f\left(\frac{1}{n}\sum_1^n x_i\right)\leqslant\frac{1}{n}\sum_1^n f(x_i). \tag{3.4}$$

证 首先明确，在所设条件下有 $y=\frac{1}{n}\sum\limits_1^n x_i\in C$. 由 f 的凸性，用归纳法易证当 $n=2^m$(m 为自然数)时(3.4)成立. 对任意的 n，找自然数 m，使 $2^m\geqslant n$. 则由已证部分，有

$$\begin{aligned}f(y)&=f\left[\frac{1}{2^m}\left(\sum_1^n x_i+(2^m-n)y\right)\right]\\&\leqslant\frac{1}{2^m}\left[\sum_1^n f(x_i)+(2^m-n)f(y)\right].\end{aligned}$$

由此即得(3.4). 引理证毕.

下面的定理把一个函数的凸性与其连续性联系起来. 其证明可参看那汤松著《实变函数论》.

定理 3.2 设 f 为定义在凸集 C 上的凸函数，而 x_0 为 C 的内点. 若存在 $\varepsilon>0$ 使

$$\inf\{f(x):\|x-x_0\|<\varepsilon\}>-\infty,$$

则 f 在 x_0 点连续.

有些作者将凸函数定义取为较强的形式：设 f 为定义在凸集 C 上的有限实函数，若对 C 中任意两点 x、y 及 $0\leqslant p\leqslant 1$，恒有

$$f(px+(1-p)y)\leqslant pf(x)+(1-p)f(y), \tag{3.5}$$

则称 f 为凸函数. 若当 $x\neq y$，且 $0<p<1$ 时，(3.5)中必有不等号成立，则称 f 为严凸的.

定理 3.3 设 f 为定义在开凸集 C 上的、在(3.3)式意义下的凸函数，则 f 在(3.5)式意义下也是凸函数的充要条件是：f 在 C 上处处连续.

本定理充分性部分证明容易. 因为由引理 3.3 易知，(3.5)式当 p 为有理数时成立，再用连续性假设，即知其对任何 $p\in[0, 1]$ 都成立. 必要性的证明基于定理 3.2，并使用归纳法，在此处从略.

需要说明的是，当凸集 C 为非开时，此定理的必要性（即：由(3.5)推得 f 的连续性）未必成立；但其充分性(即：由(3.3)及 f 的连续性推出(3.5))总是成立的.

为了避免不必要的麻烦，本书今后总是在(3.5)式的意义下使用凸函数的概念.

下面的定理对于判别函数的凸性是有用的:

定理3.4 设 f 为定义在开凸集 C 上的二阶可导的实函数，以 $A(x)$ 记 n 阶对称方阵(n 为 C 的维数)，其(i, j)元为 $\partial^2 f/\partial x_i x_j$，则

(i) 若对任何 $x\in C$, $A(x)$是非负定的，则 f 为凸函数;

(ii) 若对任何 $x\in C$, $A(x)$是正定的，则 f 为严凸函数.

证 当 $n=1$ 时，本定理的意思是: 若 $f''(x)\geqslant 0$, 对一切 x, 则 f 为凸; 若 $f''(x)>0$, 对一切 x, 则 f 为严凸. 这个事实由微分学的中值定理立即可得. 对一般的 n, 根据引理3.2, 只需证明对任意的 x、$y\in C$, 函数

$$\varphi(t)=f(tx+(1-t)y) \quad (0<t<1)$$

是凸的. 直接计算

$$\varphi''(t)=(x-y)'A(tx+(1-t)y)(x-y) \quad (0<t<1),$$

故由(i)的假定知: 对任何 $0<t<1$, $\varphi''(t)\geqslant 0$, 因而 $\varphi(t)$ 在 $0<t<1$ 是凸的. 而由 (ii) 的假定，则知在 $x\neq y$ 时，对任何 $t\in(0, 1)$, $\varphi''(t)>0$. 故 $\varphi(t)$ 在 $0<t<1$ 严凸. 这就证明了所要的结果.

三、凸函数的两条特殊性质

本段所要证明的两条性质比较特殊，在常见的著作中较少见，但对我们今后的应用却有很大的意义. 如上面已说过的，函数的凸性总是在(3.5)式的意义下，因而在开集上的凸函数总是连续的.

定理3.5 设 $f(x)$ 为定义在 R^n 上的、其下界为有限的凸函数，且 $\lim\limits_{\|x\|\to\infty} f(x)=\infty$，则存在常数 $a>0$ 及 b，使对任何 $x\in R^n$，有

$$f(x)\geqslant a\|x\|+b. \tag{3.6}$$

证 由所设条件，存在 $x_0\in R^n$，使 $f(x_0)=\inf\limits_{x} f(x)(=m)$. 取定 $k>m$. 记 $A=\{x: f(x)\leqslant k\}$, $B=\{x: f(x)=k\}$. 易见 A 为有界闭凸集，B 为 A 的边界. 记 $a=\min\{(k-m)/\|x-x_0\|: x\in B\}$,

则 $0<a<\infty$. 现在证明当 $x\overline{\in}A$ 时, 有

$$f(x)\geqslant a\|x-x_0\|+m. \tag{3.7}$$

事实上, 若(3.7)式不对, 则存在 $x^*\overline{\in}A$, 使

$$f(x^*)<a\|x^*-x_0\|+m.$$

连接 x^* 与 x_0 的线段交 B 于 x_1. 由 f 的凸性, 并注意到 a、x_0、x^*、x_1 的意义, 得出

$$\begin{aligned}f(x_1)&\leqslant\frac{\|x_1-x_0\|}{\|x^*-x_0\|}f(x^*)+\frac{\|x^*-x_1\|}{\|x^*-x_0\|}f(x_0)\\&<\frac{\|x_1-x_0\|}{\|x^*-x_0\|}[a\|x^*-x_0\|+m]+\frac{\|x^*-x_1\|}{\|x^*-x_0\|}m\\&<\frac{\|x_1-x_0\|}{\|x^*-x_0\|}(k-m+m)+\frac{\|x^*-x_1\|}{\|x^*-x_0\|}m\\&=k.\end{aligned}\tag{3.8}$$

但由 $x_1\in B$, 知 $f(x_1)=k$, 故(3.8)式出现矛盾. 这就证明了(3.7).

现取 $b=m-a\|x_0\|-a\cdot\max\{\|y-x_0\|:y\in A\}$, 则当 $x\overline{\in}A$(由(3.7))时, 有

$$a\|x\|+b\leqslant a(\|x\|-\|x_0\|)+m\leqslant a\|x-x_0\|+m\leqslant f(x);$$

而当 $x\in A$ 时, 有

$$\begin{aligned}f(x)&\geqslant m\geqslant m+a[\|x\|-\|x_0\|-\max\{\|y-x_0\|:y\in A\}]\\&=a\|x\|+b.\end{aligned}$$

这就证明了(3.6)式.

定理 3.6 设 $f(x)$ 为定义于 R^n 上的、其下界为有限的凸函数. 若存在一串点 $\{x_n\}$, 使 $\|x_n\|\to\infty$, 但 $f(x_n)$ 并不趋于 ∞, 则必存在一串趋于 ∞ 的点 $\{y_n\}$, 使

$$\lim_{n\to\infty}f(y_n)=\inf_x f(x). \tag{3.9}$$

证 记 $m=\inf\limits_x f(x)$. 找点列 $\{z_n\}$, 使 $f(z_n)\to m$. 若 $\{z_n\}$ 无界, 则从中可取出趋于 ∞ 的子列 $\{y_n\}$, 使(3.9)成立. 若 $\{z_n\}$ 有界, 则存在收敛于某点 u 的子列, 由 $f(x)$ 的连续性, 显然有 $f(u)=m$.

由对 $\{x_n\}$ 的假定, 不失普遍性, 可设 $f(x_n)\to c$ (c 为有限的).

令

$$p_n=(\|x_n-u\|-\sqrt{\|x_n-u\|})/\|x_n-u\|.$$

当 n 充分大时，有 $p_n\in(0, 1)$，故

$$f(u)\leqslant f(p_nu+(1-p_n)x_n)\leqslant p_nf(u)+(1-p_n)f(x_n).$$

令 $n\to\infty$，并注意 $p_n\to 1$，由上式得

$$\lim_{n\to\infty}f(p_nu+(1-p_n)x_n)=f(u)=m.$$

记 $y_n=p_nu+(1-p_n)x_n$，因为当 $n\to\infty$ 时，有

$$\|y_n\|\geqslant(1-p_n)\|x_n\|-p_n\|u\|=\frac{\|x_n\|}{\sqrt{\|x_n-u\|}}-p_n\|u\|\to\infty,$$

从而知 $\{y_n\}$ 即为所求的序列. 定理证毕.

四、Jensen 不等式

设 X 为 n 维随机向量，$E(X)$存在有限. 又假定存在开凸集 C，使 $P(X\in C)=1$.

定理 3.7 (Jensen 不等式)设 f 为定义于 C 上的凸函数，则 $E[f(X)]$有意义，且

$$E[f(X)]\geqslant f(E(X)). \tag{3.10}$$

又若 f 为严凸且 X 为非退化的，则(3.10)中不等号成立.

证 先注意由定理条件知 $E(X)\in C$，因而$f(E(X))$有意义. 现考虑集合

$$\widetilde{C}=\{(x, y): x\in C, y\geqslant f(x)\}. \tag{3.11}$$

利用 C 为凸集及 f 为(严)凸函数的假定，很容易证明 $\widetilde{C}$ 为(严)凸集，其边界为 $\{(x, y): x\in C, y=f(x)\}$. 故 $(E(X), f(E(X)))$ 为其一边界点. 根据定理 3.1，存在过此点的支撑平面，记其为 $b(y-f(E(X)))+a'(x-E(X))=0$，其中 a' 表示 n 维向量 a 的转置，使得对任一$(x, y)\in\widetilde{C}$，有 $b(y-f(E(X)))+a'(x-E(X))\geqslant 0$. 注意 x 为 C 的内点以及 $\widetilde{C}$ 的定义，必有 $b>0$. 因此，对任意的$(x, y)\in\widetilde{C}$，有

$$y\geqslant f(E(X))+d'(x-E(X)).$$

但当 $x\in C$ 时，$(x, f(x))\in\widetilde{C}$，从而对任何 $x\in C$，有

$$f(x) \geqslant f(E(X)) + d'(x - E(X)). \tag{3.12}$$

以 X 代(3.12)式中的 x，两边取均值，即得(3.10).

若 f 为严凸，那么 $\widetilde{C}$ 也为严凸，则由定理 3.1 知，当 $x \neq E(X)$ 时，(3.12)式取不等号. 又因 X 为非退化的，即有 $P(X \neq E(X)) > 0$，因而(3.10)中的不等号成立. 定理证毕.

§4 统计量与子 σ–域

一、统 计 量

设变量 X 的样本空间为$(\mathscr{X}, \mathscr{B}_x)$，分布族为$(P_\theta, \theta \in \Theta)$. 统计推断问题，就是要由 X 的观察值 x 去推断与 X 的分布有关的情况. 因为 X 的分布完全取决于参数 θ 的值，因此，统计推断问题也就是要根据样本 x 推断与 θ 有关的情况，例如，估计 θ 的某个函数 $g(\theta)$的值，检验某个与 θ 有关的假设等.

样本 x 是我们的原始资料，它一般表现为一大堆杂乱无章的数据，不便直接应用. 因此，无论在进行推断的准备阶段中、推断过程中，或者在推断的最后表现形式中，我们都有必要对样本进行加工整理，即算出原始资料 x 的某些特征数字 $T(x)$，它是样本 x 的函数(特别是，与参数 θ 无关)，且比较集中地体现了样本 x 中所包含的关于 θ 的信息，这就是所谓统计量. 象常用的样本均值、样本中的位数、样本方差、样本相关系数等，都是典型的统计量.

上述概念的核心就是：统计量是而且只是样本的函数. 至于函数的取值，从应用的观点看，多不外乎是一个数或向量. 但在一般理论中这一点并不必要. 对统计量的正式定义，各家的说法也略有出入. 我们认为下面的定义比较简单方便.

设$(\mathscr{X}, \mathscr{B}_x, \mathscr{P}) = (\mathscr{X}, \mathscr{B}_x, P_\theta, \theta \in \Theta)$为概率空间，$(\mathscr{T}, \mathscr{B}_\mathscr{T})$为可测空间. 任一定义在 $\mathscr{X}$ 上取值于 $\mathscr{T}$ 内的可测变换 $T = T(x)$ (即：对任何 $B \in \mathscr{B}_\mathscr{T}$，有 $T^{-1}(B) = \{x: T(x) \in B\} \in \mathscr{B}_x$) 都称为一个统计量.

在这个很一般的定义中，不仅 $\mathscr{T}$ 的选择毫无限制，而且在选

定 $\mathscr{T}$ 之后，$\mathscr{B}_{\mathscr{T}}$ 也可以任意选择．从理论上说，对同一个 $\mathscr{T}$，但 $\mathscr{B}_{\mathscr{T}}$ 选择不同，统计量的性质会受到影响．在实际问题中，$\mathscr{T}$ 往往为欧氏空间或其一非空 Borel 子集．这时，$\mathscr{B}_{\mathscr{T}}$ 一般总取为由 $\mathscr{T}$ 的一切 Borel 子集所构成．

也有这样的情况：先给定了 $\mathscr{T}$ 和 $T(x)$，而 $\mathscr{B}_{\mathscr{T}}$ 的选择就在于使 $T(x)$ 有可测性．例如：

例 4.1 设在 $\mathscr{B}_x$ 上给定了 σ–有限测度 μ，而 $\mathscr{P}\ll\mu$．记 $p(x,\ \theta)=dP_\theta(x)/d\mu$．对任何 $x\in\mathscr{X}$，确定了定义于 Θ 上的实函数 $p(x,\ \cdot)$．以 $\mathscr{T}$ 记一切这样的函数的集：$\mathscr{T}=\{p(x,\ \cdot):x\in\mathscr{X}\}$．定义 $T(x)=p(x,\ \cdot)$．定义 $\mathscr{B}_{\mathscr{T}}$ 如下：$B\in\mathscr{B}_{\mathscr{T}}$，当且仅当 $T^{-1}(B)\in\mathscr{B}_x$．不难验证 $\mathscr{B}_{\mathscr{T}}$ 为 σ–域，且 $T(x)$ 为 $(\mathscr{X},\ \mathscr{B}_x)$ 到 $(\mathscr{T},\ \mathscr{B}_{\mathscr{T}})$ 的可测变换，因而是一个统计量．

当 X 的分布为 P_θ 时，$T(X)$ 的分布为

$$P(T(X)\in B)=P_\theta(T^{-1}(B)),\ B\in\mathscr{B}_{\mathscr{T}}. \tag{4.1}$$

关系式(4.1)在 $\mathscr{B}_{\mathscr{T}}$ 上定义一个概率分布，这个分布称为 P_θ（在变换 T 之下）的导出分布，并记为 P_θ^T．如上所说，它不过就是变量 T 在参数值为 θ 时的分布而已．这样，变量 T 有了一个概率空间 $(\mathscr{T},\ \mathscr{B}_{\mathscr{T}},\ \mathscr{P}^T)$，其中 $\mathscr{P}^T=(P_\theta^T,\ \theta\in\Theta)$．这给统计量 T 的引进提供这样一种看法：引进统计量 T，等于用概率空间 $(\mathscr{T},\ \mathscr{B}_{\mathscr{T}},\ \mathscr{P}^T)$ 来代替原来的概率空间，而这是用 $T(x)$ 代替原始样本所产生的后果（统计判决问题的其他要素自然保持不变）．这个看法并非总是切合实际的，因为引进统计量 T，并未排除再引进其他（不由 T 派生出来的）统计量的可能性．然而，当统计量 T 具有所谓“充分性”（见本章 § 6）时，这种看法是合理的．

二、子 σ–域

设 T 为定义于 $(\mathscr{X},\mathscr{B}_x)$ 取值于 $(\mathscr{T},\ \mathscr{B}_{\mathscr{T}})$ 的统计量．它产生 $\mathscr{B}_x$ 的一个子 σ–域：

$$T^{-1}(\mathscr{B}_{\mathscr{T}})=\{A:A=T^{-1}(B),\ B\in\mathscr{B}_{\mathscr{T}}\},$$

这个 σ–域称为统计量 T 的导出 σ–域．显然，有 $T^{-1}(\mathscr{B}_{\mathscr{T}})\subset\mathscr{B}_x$．故

任何统计量 T 的导出 σ-域都是 $\mathscr{B}_x$ 的子 σ-域.

由于子 σ-域是由统计量决定的，这就引起人们一种想法：能否用子 σ-域来代替统计量？如果可行，这样做在理论上的好处，一是不需脱离原来的空间 $(\mathscr{X}, \mathscr{B}_x)$ 而去引进另一个空间 $(\mathscr{T}, \mathscr{B}_{\mathscr{T}})$；一是从某种意义上说，子 σ-域更深刻地反映了统计量的本质.

例 4.2 $(\mathscr{X}, \mathscr{B}_x)=(R^1, \mathscr{B}^1)$, $\mathscr{T}=\{t: t\geqslant 0\}$, $\mathscr{B}_{\mathscr{T}}$ 为 $\mathscr{T}$ 的 Borel 子集构成的 σ-域. 考虑

$$T_1(x)=x^2,\ T_2(x)=|x|.$$

这两个统计量表面上完全不同，实质上二者是等价的（由其一个可决定另一个），且都在于把两个关于原点对称的样本点简化为一个. 如果从子 σ-域看，则不难见到：二者所导出的子 σ-域完全一致，都是由 R^1 的一切对称 Borel 子集所构成的 σ-域.

在比较简单的情况下，由子 σ-域本身直接就可看出对原样本 x 作了怎样的简化. 就上例来说，引进一个由 R^1、ϕ、$(-\infty, 0)$ 和 $[0, \infty)$ 四个集构成的子 σ-域 $\mathscr{F}$. 这个子 σ-域相当于对原样本 x 作如下的简化：所有 $\geqslant 0$ 的样本简化为一个值 a；所有 <0 的样本简化为另一个值 $b\neq a$. 这就是说，$\mathscr{F}$ 是下面这一类统计量

$$T(x)=\begin{cases} a, & x\geqslant 0; \\ b, & x<0 \end{cases} \quad (a\neq b)$$

的公共导出 σ-域. 但是，在子 σ-域不是那么简单的情况，就不易看出它对样本究竟作了怎样的简化.

至于在理论上说，究竟子 σ-域与统计量能否互相代替，需要从以下几点上来考虑.

(1) 用一个能办到的事，而用另一个能否办到（这问题的确切含义见下文可知）？

(2) 二者能否彼此决定？

(3) 在两种概念（即统计量和子 σ-域）下所产生的结果能否平行地彼此推过去？

为了讨论前两个问题，我们要引用下面两个很重要的定理. 其证明将在本节末尾给出.

定理 4.1 沿用前面的记号，设 f 为由$(\mathscr{X}, \mathscr{B}_x)$到$(R^m, \mathscr{B}^m)$的可测变换，则当 f 为 $T^{-1}(\mathscr{B}_{\mathscr{T}})$ 可测时，必存在由$(\mathscr{T}, \mathscr{B}_{\mathscr{T}})$ 到$(R^m, \mathscr{B}^m)$的可测变换 φ，使对任何 $x \in \mathscr{X}$，有

$$f(x)=\varphi[T(x)]. \tag{4.2}$$

反过来，若存在这样的 φ 使(4.2)成立，则 f 必为 $T^{-1}(\mathscr{B}_{\mathscr{T}})$可测.

定理 4.2 设 μ 为 $(\mathscr{X}, \mathscr{B}_x)$ 上的测度，而 μ^T 为 μ 在变换 T 之下的导出测度，则对任何 $\mathscr{B}_{\mathscr{T}}$-可测函数 g 和 $B \in \mathscr{B}_{\mathscr{T}}$，有

$$\int_{T^{-1}(B)} g(T(x))d\mu(x)=\int_B g(t)d\mu^T(t). \tag{4.3}$$

更确切地说，当(4.3)的一边有意义时，另一边也有意义，且二者相等.

根据这两个定理就可以对上面提的第一个问题基本上给以肯定的回答：若使用统计量 T，则表示日后在进行统计推断时，我们只使用基于 T 的(可测)函数，这种函数是 $T^{-1}(\mathscr{B}_{\mathscr{T}})$ 可测的，反之亦然. 因此，如把"使用子 σ-域 $\mathscr{F}$"理解为"日后在进行统计推断时，只使用 $\mathscr{F}$-可测的函数"，则根据定理 4.1，使用 T 能作到的事，使用 $T^{-1}(\mathscr{B}_{\mathscr{T}})$也能作到，反之亦然. 定理 4.2 保证了二者在积分运算上的等价性.

关于第二个问题，统计量 T 决定其导出的子 σ-域，这一部分是清楚的. 反面的问题，又可分为两部分：其一是：任给一子 σ-域 $\mathscr{F} \subset \mathscr{B}_x$，是否存在统计量 T，使 $T^{-1}(\mathscr{B}_{\mathscr{T}})=\mathscr{F}$？如果对$(\mathscr{T}, \mathscr{B}_{\mathscr{T}})$不作任何限制，这个问题的回答显然是肯定的，因为只需取 $\mathscr{T}=\mathscr{X}$, $\mathscr{B}_{\mathscr{T}}=\mathscr{F}$, 而 $T(x)=x$. 但从较为实用的角度来考虑，往往需对统计量的值域空间加上一定的限制，特别是，假定它是欧氏的，在这种情况下，就不一定每个子 σ-域 $\mathscr{F}$ 都能由某个统计量导出. 下面就是一个例子(即例 4.3)，验证是容易的，我们把它作为练习留给读者.

例 4.3 设$(\mathscr{X}, \mathscr{B}_x)=(R^1, \mathscr{B}^1)$，要求统计量的值域空间为$(R^1, \mathscr{B}^1)$. 以 $\mathscr{F}$ 记一切"本身或其余集为可列"的集构成的 σ-域(空集、有限集都称作可列)，则 $\mathscr{F}$ 不能由任何统计量导出.

第二个问题的另一方面是：如果一个子 σ-域 $\mathscr{F}\subset\mathscr{B}_x$ 能够由某一统计量导出，那么，是否会有一些"本质上不一样"的统计量都导出这个 σ-域？如果回答是肯定的，则我们就没有理由认为，统计量与子 σ-域能彼此决定. 对这个问题，必须明确"本质上不一样"是什么意思. 为此，我们不妨暂时引进如下的概念：称两个统计量

$$T:(\mathscr{X},\ \mathscr{B}_x)\to(\mathscr{T},\ \mathscr{B}_T);\ S:(\mathscr{X},\ \mathscr{B}_x)\to(\mathscr{S},\ \mathscr{B}_S)$$

等价，若存在 $(\mathscr{T},\ \mathscr{B}_T)$ 到 $(\mathscr{S},\ \mathscr{B}_S)$ 的可测变换 φ，以及 $(\mathscr{S},\ \mathscr{B}_S)$ 到 $(\mathscr{T},\mathscr{B}_T)$ 的可测变换 ψ，使对一切 $x\in\mathscr{X}$，有

$$s(x)=\varphi[T(x)],\ T(x)=\psi[s(x)].\tag{4.4}$$

不难理解，如果两个统计量 T 和 S 等价，那么，即使其具体形式可能有很大差异，但它们"在实质上"是一回事. 因为(4.4)说明了 T 和 S 可以彼此决定：知道其一可以算出其二. 这样，我们自然可把两个不等价的统计量称为是"本质上不一样"的. 故问题归结为：同一个子 σ-域能否由本质上不一样的统计量导出. 由定理 4.1 可立即推出：若限制统计量的值域空间为欧氏的，则这个问题的回答为"否". 但是，如果对统计量的值域空间不加限制，则问题就大为复杂化了，不能给出一般的回答.

至于第三个问题，由于更加复杂且不十分确定，就更难于给出简单的回答了. 由于与本书今后的关系不大，我们就不在这一点上作深入讨论了.

综合以上讨论的结果，我们可以认为：使用子 σ-域来讨论统计量是一个有用的工具，但用它来代替统计量则是不妥的. 这个问题在文献中花费了不少篇幅，总的来说意义不大. 本书作者的观点是赞成以统计量作为根本出发点而只是把子 σ-域作为一个附属物. 无论如何，统计量所具有的直观性是子 σ-域所不及的.

三、定理 4.1 和 4.2 的证明

定理 4.1 的证明 定理后一部分的证明是显然的. 为证前一部分，取 $m=1$ 的情况来讨论（对一般的 m，下面的证明完全适

用). 若 $f(x)=I_A(x)$, 其中 $A\in T^{-1}(\mathscr{B}_T)$, 则适合(4.2)的 φ 可取为 $\varphi(t)=I_B(t)$, 这里 $B\in\mathscr{B}_T$, 而 $A=T^{-1}(B)$. 因此, 对形如

$$f(x)=\sum_{i=1}^{c} a_i I_{A_i}(x) \quad (a_i\text{ 为常数}, A_i\in T^{-1}(\mathscr{B}_T),$$

$$i=1, \cdots, c, \quad c\text{ 为任何自然数}) \qquad (4.5)$$

的函数, 定理结论也成立. 现设 f 为任一 $T^{-1}(\mathscr{B}_T)$-可测函数, 则存在一半 $\{f_n\}$, 使对任何 $x\in\mathscr{X}$, 有

$$\lim_{n\to\infty} f_n(x)=f(x),$$

其中每个 f_n 有(4.5)的形式. 因此对每一 n, 存在 $\mathscr{B}_T$-可测的 φ_n, 使对任何 $x\in\mathscr{X}$, 有

$$f_n(x)=\varphi_n(T(x)) \quad (n=1, 2, \cdots).$$

现令
$$\varphi(t)=\begin{cases}\lim\limits_{n\to\infty}\varphi_n(t), & \text{若此极限存在有限;}\\ 0, & \text{其它情况,}\end{cases}$$

则 φ 为 $\mathscr{B}_T$-可测, 且对任何 x, 有

$$\lim_{n\to\infty}\varphi_n(T(x))=\lim_{n\to\infty}f_n(x)=f(x).$$

因而 $\lim\limits_{n\to\infty}\varphi_n(T(x))$ 存在, 故依 φ 的定义, 有 $\varphi(T(x))=f(x)$. 定理证毕.

定理 4.2 的证明 不难看到, 只需考虑 $B=\mathscr{T}$ 的情况. 又易见若 $g(t)=I_B(t)(B\in\mathscr{B}_T)$时, (4.3)式成立(这就是 μ^T 的定义). 于是用测度论的标准证法, 依次推到非负简单可测函数、非负可测函数及一般可测函数的情况.

§5 条件期望与条件概率分布

条件期望和条件概率分布是概率论基础中的重要内容. 在一般的较深的概率论著作中有仔细的讨论(例如见[1]). 但是, 由于这些内容对本书极为重要, 所以我们还是安排这一节, 以适合于统计需要的形式, 把这个理论中的基本事实作一简明的叙述. 在这方面, [2] 中 (见第二章)的叙述方式很适合我们的要求, 但过于

简略.

一、 条件期望与条件概率

1. 条件期望的定义(对统计量) 正如我们在 §1 中所指出的,统计问题涉及到一个分布族. 但就本节问题而论,并不需要考虑参数 θ 变化的情况, 因此不妨设只有一个分布. 这样, 设变量 X 的概率空间为$(\mathscr{X}, \mathscr{B}_x, P)$.

设 $f(X)$为一维随机变量[即: f 为$(\mathscr{X}, \mathscr{B}_x)\to(R^1, \mathscr{B}^1)$的可测变换],其均值存在有限,即

$$E(|f(X)|)=\int_{\mathscr{X}}|f(x)|dP(x)<\infty, \tag{5.1}$$

又设 $T:(\mathscr{X},\mathscr{B}_x)\to(\mathscr{T},\mathscr{B}_{\mathscr{T}})$为一统计量,我们要考虑在给定 T 的条件下 $f(X)$ 的条件期望的定义问题. 这个问题的直观背景是很清楚的. 设想这样一种情况. 在对 X 观察后得出样本 x. 但是并不知道 x, 因为观察者通过计算统计量 $t=T(x)$的值后, 丢掉了 x. 你所知道的只是 $T(x)=t$. 这时, 站在你所处的地位去看, 在计算有关的期望或概率时, 合理的作法是考虑它们在给定 $T(x)=t$ 的条件下的"条件值".

我们在初等概率中曾碰到过条件期望值的计算. 不同的是, 那里给出的条件总有正的概率. 例如, 若不是给定 $T(x)=t$ 而是给定 $T\in B(B\in\mathscr{B}_T)$, 且 $P(T(X)\in B)>0$, 则有

$$E[f(X)|T\in B]=\int_{T^{-1}(B)}f(x)dP(x)/P(T\in B). \tag{5.2}$$

但一般说来, 对一个 t 值 $P(T=t)=0$, 因而按(5.2)定义 $E[f(X)|T=t]$有困难. 解决这个困难的思路从整体性质上去考虑条件期望应当如何, 而不是对逐个的 t 定义 $E[f(X)|T=t]$. 举个浅显的例子: 要逐点定出一根细棒的线密度 $\rho(x)$ 可能有困难, 但 $\rho(x)$ 的整体性质是: 对这棒上的任两点 $a<b$, $\int_a^b\rho(x)\,dx$ 应等于此棒在 $[a, b]$ 段的质量, 这个性质可作为定义线密度函数 $\rho(x)$的基础, 而不必对每个 x 去考虑极限

$$\lim_{h\downarrow 0,\ k\downarrow 0}[\text{棒在}[x-h,\ x+k]\text{段的质量}\div(h+k)].$$

我们把这个思想用于条件期望 $E[f(X)|T=t]=E[f(X)|t]$ 的定义. 显然, 这个条件期望是 t 的函数, 暂记为 $m(t)$. 其次, 若把 $m(t)$ 在一集合上再求一次平均, 其结果应当等于 $f(x)$ 在这个集合上的平均, 因为这相当于在求 $f(x)$ 的平均时分两步走: 先固定 t 求平均, 再对 t 求平均. 它当然应与一次求出的平均一致. 这样, 我们得到 $m(t)$ 应满足的关系式如下:

$$\int_B m(t)dP^T(t)/P^T(B)=\int_{T^{-1}(B)} f(x)dP(x)/P(T^{-1}(B)), \tag{5.3}$$

对任何 $B\in\mathscr{B}_T$, $P(T^{-1}(B))>0$. 这里 P^T 为 T 的导出测度, 即 $P^T(B)=P(T^{-1}(B))$. 读者很容易看出: (5.3)式是上面我们指出的"两步平均等于一步平均"这个思想的公式体现. 由(5.3)得

$$\int_B m(t)dP^T(t)=\int_{T^{-1}(B)} f(x)\,dP(x),\ B\in\mathscr{B}_T. \tag{5.4}$$

当 $P(T^{-1}(B))=0$ 时, (5.4)当然成立.

由上面讨论的启发, 我们引进下面的正式定义.

定义 5.1 在前面的记号下, 所谓"在给定 $T(x)=t$ 的条件下, $f(X)$ 的条件期望", 是指任一满足条件(5.4)的、$\mathscr{B}_{\mathscr{T}}$-可测的函数 $m(t)$.

这个定义没有指明 $m(t)$ 的存在唯一性. 因此, 下面的引理是定义 5.1 的必要补充.

引理 5.1 设(5.1)成立, 则满足(5.4)式的 $m(t)$ 必存在, 且在 $a.s.\ P^T$ 的意义下是唯一的.

证 在 $T^{-1}(\mathscr{B}_T)$ 上定义

$$\nu(A)=\int_A f(x)dP(x),\ A\in T^{-1}(\mathscr{B}_T). \tag{5.5}$$

在条件(5.1)之下, ν 显然是一个有限的符号测度, 且 $\nu\ll P$. 于是依 Radom-Nikodym 定理, 存在 $T^{-1}(\mathscr{B}_T)$-可测的函数 $\tilde{f}(x)$, 使

$$\nu(A)=\int_A \tilde{f}(x)dP(x),\ A\in T^{-1}(\mathscr{B}_T). \tag{5.6}$$

根据定理4.1及(4.2)，存在$\mathscr{B}_T$-可测函数$m(t)$，使$\tilde{f}(x)=m(T(x))$，对任何$x\in\mathscr{X}$，且

$$\int_A \tilde{f}(x)dP(x)=\int_B m(t)dP^T(t),\ B\in\mathscr{B}_T,\ A=T^{-1}(B). \tag{5.7}$$

由(5.5)～(5.7)得(5.4)，这就证明了$m(t)$的存在性．又若存在$m_1(t)$、$m_2(t)$都满足(5.4)，并记$m^*=m_1-m_2$，对任何$B\in\mathscr{B}_T$，将有

$$\int_B m^*(t)\,dP^T(t)=0.$$

由于$m^*(t)$为$\mathscr{B}_T$-可测，有$P^T(m^*(T)=0)=1$．这就证明了引理中的唯一性部分．

除非有特殊声明，以后在谈到$E[f(X)|t]$时，总是指某一确定的满足(5.4)的$m(t)$．又，记号$E[f(X)|T]$是指$m(T)=E[f(X)|t]_{t=T}$．必要时，为了清楚起见，也将$E[f(X)|t]$写成$E[f(X)|T=t]$．

2．条件期望的定义（对子σ-域）　统计量T导出子σ-域$T^{-1}(\mathscr{B}_T)$．在§4中我们讨论过统计量与子σ-域的关系，并提到有时使用子σ-域是方便的．因此自然地产生怎样在给定子σ-域的条件下定义条件期望的问题．

给定$\mathscr{B}_x$的子σ-域$\mathscr{F}$以及满足条件(5.1)的$f(x)$．“在给定$\mathscr{F}$的条件下$f(X)$的条件期望”$E[f(X)|\mathscr{F}]$定义为任一满足条件

$$\int_A \tilde{f}(x)dP(x)=\int_A f(x)dP(x),\ A\in\mathscr{F} \tag{5.8}$$

的$\mathscr{F}$-可测函数$\tilde{f}$．类似于引理5.1的结果及其证法，在此显然仍适用．因此，$E[f(X)|\mathscr{F}]$存在且在$a.s.$ $(\mathscr{F},\ P)$意义下唯一．

这个定义的直观意义可用几种方法解释：首先，若$\mathscr{F}$为由集$A_1,\cdots,A_r$生成的σ-域，其中$A_i\in\mathscr{B}_x$两两不相交，且$\bigcup_{i=1}^r A_i=\mathscr{X}$，则$E[f(x)|\mathscr{F}]$转化为初等概率论中所熟知的情况，即：对$x\in A_i$，(5.8)中的$\tilde{f}(x)$为

$$\bar{f}(x)=E[f(X)|A_i]=\frac{1}{P(A_i)}\int_{A_i}f(x)dP(x),\ x\in A_i.$$

另一个解释如下：设想 $\mathscr{F}=T^{-1}(\mathscr{B}_T)$. $\bar{f}(x)=E[f(X)|\mathscr{F}]$既然为 $T^{-1}(\mathscr{B}_T)$可测，故可表为 $\varphi(T(x))$ 的形状．因而由定理4.2，有

$$\int_A f(x)dP(x)=\int_B\varphi(t)dP^T(t),\ B\in\mathscr{B}_T,\ A=T^{-1}(B).$$

这表明：$\varphi(t)$为条件期望 $E[f(X)|t]$. 因此，这样通过子 σ-域定义的条件期望不过是通过统计量定义的条件期望的另一种表现形式.

例 5.1 设 $X_1,\ X_2,\ \cdots,\ X_n$ 为 $iid.$ 随机变量，X_1 的分布为 F. $X=(X_1,\ \cdots,\ X_n)$，样本空间为$(\mathscr{X},\ \mathscr{B}_x)=(R^n,\ \mathscr{B}^n)$. X 的分布为 $P=F\times\cdots\times F$. 引进可测空间$(\mathscr{T},\ \mathscr{B}_T)$，这里

$$\mathscr{T}=\{t:t=(t_1,\ \cdots,\ t_n),\ t_1\leqslant t_2\leqslant\cdots\leqslant t_n\},$$

而 $\mathscr{B}_T$ 为 T 的一切 Borel 子集构成的 σ-域．定义统计量 $T(X)=(X_{(1)},\ \cdots,\ X_{(n)})$，这里 $X_{(1)}\leqslant X_{(2)}\leqslant\cdots\leqslant X_{(n)}$ 为 $X_1,\ \cdots,\ X_n$ 的从小到大的排列，即次序统计量．设 $f(x)$为 n 元 Borel 可测函数，$\int_{R^n}|f(x)|dP(x)<\infty$，则有

$$E[f(X)|T(X)=t]=\frac{1}{n!}\sum_{(i_1,\cdots,i_n)}f(t_{i_1},\ \cdots,\ t_{i_n}).\qquad(5.9)$$

这里求和范围是对$(1,\ 2,\ \cdots,\ n)$的全部置换$(i_1,\cdots,i_n)$，而 $t=(t_1,\ \cdots,\ t_n)$，$t_1\leqslant t_2\leqslant\cdots\leqslant t_n$. 事实上，由 f 的 Borel 可测性及(5.9)式右边的形状，即知若将(5.9)的右边记为 $m(t)$，则 $m(t)$为 $\mathscr{B}_T$-可测．又对任何 $B\in\mathscr{B}_T$，有

$$\int_B m(t)dP^T(t)=\frac{1}{n!}\sum_{(i_1,\cdots,i_n)}\int_A f(x_{i_1},\ \cdots,\ x_{i_n})dF(x_1)\cdots dF(x_n),\qquad(5.10)$$

此处 $A=T^{-1}(B)$，故 A 为对称的 Borel 集，因此

$$\int_A f(x_{i_1},\ \cdots,\ x_{i_n})dF(x_1)\cdots dF(x_n)$$
$$=\int_A f(x_1,\ \cdots,\ x_n)dF(x_1)\cdots dF(x_n),$$

对任何置换$(i_1, \cdots, i_n)$. 由此及(5.10), 得

$$\int_B m(t)dP^T(t)=\int_A f(x_1, \cdots, x_n)dF(x_1)\cdots dF(x_n).$$

这就证明了(5.9)式. 注意(5.9)的右边与分布F无关.

3. 条件期望的性质　我们从统计量的角度来讨论, 因为子σ-域的情况完全相似. 我们把条件期望的性质逐条开列如下, 然后再来证明:

(i) $E(c|t)=c$, $a.s.P^T$, 这里c为常数.

(ii) 若$f(x)\leqslant g(x)$, 对x $a.s.P$成立, 则$E[f(X)|t]\leqslant E[g(X)|t]$, $a.s.P^T$.

由此性质可知, 若f非负, 则$E[f(X)|t]$也非负$(a.s.P^T)$, 又若$a\leqslant f\leqslant b$(a, b为常数), 则$a\leqslant E[f(X)|t]\leqslant b$ $(a.s.P^T)$.

(iii) $E\{E(f(X)|T)\}=E[f(X)]$.

这条性质的直观意义是"两步平均等于一步平均", 这是一条常用的重要性质.

(iv) 若a、b为常数, 则

$$E[af(X)+bg(X)|t]$$
$$=aE[f(X)|t]+bE[g(X)|t] \quad a.s.P^T.$$

这个性质显然可推广到有限项的和.

在以上这几条性质中, 都假定了$f(X)$、$g(X)$等的均值存在有限. 这几个性质的证明, 很容易直接从定义得出, 因此把它作为练习留给读者.

在下面几条性质中, 出现的$f(X)$、$f_n(X)$、$g(X)$等, 都假定均值存在有限.

(v) **单调收敛定理**　设$g(x)\leqslant f_1(x)\leqslant f_2(x)\leqslant\cdots\rightarrow f(x)$, 对$x$ $a.s.P$成立, 则

$$\lim_{n\to\infty}E[f_n(X)|t]=E[f(X)|t] \quad a.s.P^T. \tag{5.11}$$

(vi) **控制收敛定理**　设$|f_n(x)|\leqslant g(x)$ $(n=1, 2, \cdots)$, 对x $a.s.P$成立, 且$\lim\limits_{n\to\infty}f_n(x)=f(x)$, $a.s.P$, 则(5.11)式成立.

(vii) 设 $h(t)$ 为 $\mathscr{B}_T$-可测函数,且

$$E[|f(X)|]<\infty, E[|f(X)h(T(X))|]<\infty,$$

则

$$E[f(X)h(T(X))|t]=h(t)E[f(X)|t] \quad a.s.P^T. \tag{5.12}$$

现在我们来证明(v)～(vii):

(v)的证明 由假定及性质(ii)知,存在 $\mathscr{B}_T$-可测的函数 φ,使

$$E[g(X)|t]\leqslant E[f_n(X)|t]\uparrow\varphi(t), \quad a.s.P^T.$$

由积分的单调收敛定理知,对任何 $B\in\mathscr{B}_T$,有

$$\int_B\varphi(t)dP^T(t)=\lim_{n\to\infty}\int_B E[f_n(X)|t]dP^T(t)$$

$$=\lim_{n\to\infty}\int_A f_n(x)dP(x)=\int_A f(x)dP(x), \tag{5.13}$$

此处 $A=T^{-1}(B)$. (5.13)式证明了 $\varphi(t)=E[f(X)|t]$.

(vi)的证明 记 $h_n(x)=\sup_{k>n}f_k(x)$, $l_n(x)=\inf_{k\geqslant n}f_k(x)$,则 $g(x)\geqslant h_1(x)\geqslant h_2(x)\geqslant\cdots\to f(x)$, $a.a.P$, 以及 $-g(x)\leqslant l_1(x)\leqslant l_2(x)\leqslant\cdots\to f(x)$, $a.s.P$,根据刚才证明的(v)(及其在单调下降情况的显然推广),可知

$$E[h_n(X)|t]\to E[f(X)|t], \quad a.s.P^T, \tag{5.14}$$

$$E[l_n(X)|t]\to E[f(X)|t], \quad a.s.P^T. \tag{5.15}$$

再利用 $l_n(x)\leqslant f_n(x)\leqslant h_n(x)$ 及性质(ii),由(5.14)及(5.15)得(5.11).

(vii)的证明 若 $h(t)=I_B(t)$, $B\in\mathscr{B}_T$, 则(5.12)显然成立.于是,由性质(iv)可知,当 $h(t)$ 为有限简单 $\mathscr{B}_T$-可测函数时,(5.12)式成立.对一般情况,作一串有限简单 $\mathscr{B}_T$-可测函数 $\{h_n(t)\}$ 使 $|h_n(t)|\leqslant|h(t)|$,且 $\lim_{n\to\infty}h_n(t)=h(t)$ 对任何 $t\in\mathscr{T}$,再利用(vi)即得.

性质(vii)有一个简单的直观解释:在固定 T 时,$h(T)$ 成为一个常数,因此可提到期望符号 E 的外面.特别有

$$E[h(T(X))|t]=h(t), a.s. P^T. \tag{5.16}$$

由此得到一个在直观上显然的结果. 设 $h(t)$ 为 $\mathscr{B}_T$-可测, $S=h(T)$, 则

$$E\{E[f(X)|S]|T\}=E[f(X)|S], \quad a.s. P^T. \tag{5.17}$$

事实上, $E[f(X)|S]$ 为 S 的 Borel 可测函数, 故更为 T 的 Borel 可测函数, 因此(5.17)成为(5.16)的特殊情况. 另一方面, 也有

$$E\{E[f(X)|T]|S\}=E[f(X)|S], \quad a.s. P^S. \tag{5.18}$$

这个性质很容易由条件期望的定义直接证明, 我们把它作为练习留给读者.

在子 σ-域的情况, 有如下相应的性质:

(viii) 设 $\mathscr{F}_1$ 与 $\mathscr{F}_2$ 都是 $\mathscr{B}_x$ 的子 σ-域, $\mathscr{F}_2\subset\mathscr{F}_1$, 则

$$\begin{aligned}E\{E[f(X)|\mathscr{F}_1]|\mathscr{F}_2\}&=E[E[f(X)|\mathscr{F}_2]|\mathscr{F}_1\}\\&=E[f(X)|\mathscr{F}_2], \quad a.s. P.\end{aligned}$$

证明容易由定义直接得出.

4. 条件概率　正如通常的(无条件的)概率是数学期望的特例那样, 条件概率也是条件期望的特例.

定义 5.2　沿用前面记号. 设 $A\in\mathscr{B}_x$, 则"在给定 $T(X)=t$ 的条件下, 事件 A 的条件概率" $P(A|t)$ 定义为 $E[I_A(X)|t]$.

类似地定义 $P(A|\mathscr{F})$, $\mathscr{F}$ 为 $\mathscr{B}_x$ 的子 σ-域. 由于条件概率是条件期望的特例, 不仅其存在唯一性不需另行论证, 其性质也作为条件期望性质的特例可立即写出来. 例如:

(i) $0\leqslant P(A|t)\leqslant 1$, $a.s. P^T$;

(ii) $E[P(A|T)]=P(A)$;

(iii) 若 $\lim\limits_{n\to\infty}A_n=A$, 则 $\lim\limits_{n\to\infty}P(A_n|t)=P(A|t)$, $a.s. P^T$;

(iv) 若 $A_1, A_2, \cdots$ 两两不相交, $A=\bigcup\limits_{n=1}^{\infty}A_n$, 则

$$P(A|t)=\sum_{n=1}^{\infty}P(A_n|t), \quad a.s. P^T. \tag{5.19}$$

此处性质(i)、(ii)、(iii)分别是条件期望性质(ii)、(iii)、(vi)的特例, 而此处性质(iv)是(iii)的特例. 值得注意的是: (5.19)式中的

例外集合与 $\{A_n\}$ 有关.

二、条件概率分布

1. 定义和存在性　根据定义(5.2),条件概率 $P(A|t)$ 可以看作是一个定义于 $\mathscr{B}_x\times\mathscr{T}$、取值于[0, 1]的函数，当 $A\in\mathscr{B}_x$ 固定时,$P(A|t)$是 $\mathscr{B}_T$-可测函数,且满足(5.19).

在前面我们曾特别指出:(5.19)式的例外集与 $\{A_n\}$ 有关. 这样一来, 对任意固定的 $t\in\mathscr{T}$ 及两两不相交的 $A_n\in\mathscr{B}_x$, 我们不能断言 $P\left(\bigcup_1^\infty A_n|t\right)=\sum_{n=1}^\infty P(A_n|t)$, 因为这个 t 可能正好就在 $\{A_n\}$ 的例外集中. 这样一来, 当 t 固定时, 没有理由断言 $P(A|t)$ 是 $\mathscr{B}_x$ 上的概率分布. 但是不言而喻, 如果能有 $P(A|t)$的一种取法(因为 $P(A|t)$ 对每一固定的 A 都是 $a.s.\ P^T$ 唯一确定的, 故有选择的余地),使当 t 固定时,$P(A|t)$为 $\mathscr{B}_x$ 上的概率分布,则在理论上将有莫大的好处, 而且也使条件期望与条件概率的关系更显得自然. 这个考虑引导到下面的重要定义.

定义 5.3　定义在 $\mathscr{B}_x\times\mathscr{T}$ 上的函数 $P(A, t)$ 称为是给定 T 的条件下的正则条件概率分布(简称为条件概率分布), 如果它满足以下两个条件:

1° 在给定 $A\in\mathscr{B}_x$ 时, $P(A, t)$ 作为 t 的函数, 是 $\mathscr{B}_T$-可测的,且 $P(A, t)=P(A|t)$, $a.s. P^T$.

2° 在给定 $t\in\mathscr{T}$ 时,$P(A, t)$作为 A 的函数,为 $\mathscr{B}_x$ 上的概率测度.

如果 $P_1(A, t)$和 $P_2(A, t)$都是条件概率分布函数, 则由条件 1° 得 $P_1(A, t)=P_2(A, t)$, $a.s. P^T$,对任何固定的 $A\in\mathscr{B}_x$(例外集可与 A 有关). 在这个意义上可以说条件概率分布有唯一性. 但其存在性如何, 则与样本空间$(\mathscr{X}, \mathscr{B}_x)$有关. 存在这样的例子, 其中条件概率分布函数不存在(见[1]). 但是, 有下面的重要定理:

定理 5.1　若$(\mathscr{X}, \mathscr{B}_x)$为欧氏的, 则条件概率分布函数存在.

证 为确定计，设$(\mathscr{X},\ \mathscr{B}_x)=(R^1,\ \mathscr{B}^1)$. 一般情况证明完全相似.

对任意实数x，取定一条件概率$P((-\infty,\ x]\mid t)$，记为$F(x,\ t)$. 将全部有理数排列成r_1、r_2、$\cdots$. 则对任何i、j，若$r_i<r_j$，存在$N_{ij}\in\mathscr{B}_T$，使$P^T(N_{ij})=0$，当$t\overline{\in}N_{ij}$时，

$$F(r_i,\ t)\leqslant F(r_j,\ t).$$

又由条件概率的性质(iii)，存在P^T-零集$N_1,\ N_2,\ \cdots$及N'、N''，使

$$\lim_{n\to\infty}F\left(r_i+\frac{1}{n},\ t\right)=F(r_i,\ t)\quad (\text{当 } t\overline{\in}N_i,\ i=1,\ 2,\ \cdots);$$

$$\lim_{n\to\infty}F(n,\ t)=1\quad (\text{当 } t\overline{\in}N');$$

$$\lim_{n\to\infty}F(-n,\ t)=0\quad (\text{当 } t\overline{\in}N'').$$

记$N=(\cup_{ij}N_{ij})\cup(\cup_iN_i)\cup N'\cup N''$. N仍为P^T-零集. 当$t\overline{\in}N$时，$F(x,\ t)$在x取有理值时，具有一维分布函数的全部特征. 现对$t\overline{\in}N$定义

$$F^*(x,\ t)=\lim_{x'\downarrow x}F(x',\ t),$$

此处x'只取有理数值. 又任取$t_0\overline{\in}N$，而当$t\in N$时，定义$F^*(x,t)=F^*(x,\ t_0)$. 这样，就定义了$\mathscr{X}\times\mathscr{T}$上的函数$F^*(x,\ t)$，它具有如下的性质：

1° 当固定x时，$F^*(x,\ t)$作为t的函数，为$\mathscr{B}_T$-可测；

2° 当固定t时，$F^*(x,\ t)$作为x的函数，是一个一维分布函数.

根据2°，由$F^*(x,\ t)$可以产生一个定义在$\mathscr{B}_x\times\mathscr{T}$上的函数$P(A,\ t)$，它作为$t$的函数为$\mathscr{B}_T$-可测，而作为$A$的函数为$\mathscr{B}_x$上的概率测度. 为了证明本定理，只需证明：对任何$A\in\mathscr{B}_x$，有

$$P(A,\ t)=P(A\mid t),\quad a.s.P^T. \tag{5.20}$$

以$\mathscr{F}$记一切满足(5.20)的A的集类，则由$P(A,\ t)$的定义过程知，$\mathscr{F}$包含一切有理区间，又由条件概率的性质(iii)，易知$\mathscr{F}$为一单调类. 于是根据周知的定理（见[1]），$\mathscr{F}$包含由直线上一切

有理区间所生成的σ-域，即$\mathscr{B}^1$. 这就完成了定理的证明.

考虑到在统计问题中，样本空间几乎全是欧氏的，这定理的意义就可想而知了.

2. 条件概率分布的性质　仍沿用前面的记号，而且碰到求期望或条件期望时，总假定它们存在有限.

定理 5.2　若$P(A, t)$为条件概率分布，则

$$E[f(X)|t]=\int_{\mathscr{X}} f(x)P(dx, t), \quad a.s. P^T. \tag{5.21}$$

证　由$P(A, t)$的定义，(5.21)式当$f(x)=I_A(x)$时成立. 于是由测度论的标准证法，即得(5.21)式对一般的f成立.

这个定理说明：在条件概率分布存在的情况下，条件期望等于变量(即$f(x)$)对条件概率分布的积分，恰如在无条件下，数学期望与概率分布之间的关系.

设条件概率分布$P(A, t)$存在，又设$\mathscr{T}$的每个单点集$\{t\}\in\mathscr{B}_T$，则$A_t=\{x: T(x)=t\}\in\mathscr{B}_x$.

根据条件概率的直观含义，在给定了$T=t$时，有理由期望全部概率集中在集A_t上，即$P(A_t, t)=1$. 关于这个问题有如下的结果：

定理 5.3　设条件概率分布$P(A, t)$存在(例如，在$(\mathscr{X}, \mathscr{B}_x)$为欧氏时)，且$(\mathscr{T}, \mathscr{B}_T)$为欧氏的，则存在$P^T$-零集$N$，使

$$P(A_t, t)=1, \text{当 } t\bar{\in} N. \tag{5.22}$$

证　为确定计，设$(\mathscr{T}, \mathscr{B}_T)=(R^1, \mathscr{B}^1)$. 一般情况的证明完全相似.

先直接根据定义验证对于任何$B\in\mathscr{B}_T$，有

$$P(T^{-1}(B), t)=I_B(t), \quad a.s. P^T. \tag{5.23}$$

事实上，$I_B(t)$为$\mathscr{B}_T$-可测函数，且对任何$B_1\in\mathscr{B}_T$，有

$$\begin{aligned}\int_{B_1} I_B(t)\,dP^T(t)&=P^T(B\cap B_1)=P(T^{-1}(B\cap B_1))\\&=P(T^{-1}(B)\cap T^{-1}(B_1))\\&=\int_{T^{-1}(B_1)} I_{T^{-1}(B)}(x)dP(x).\end{aligned}$$

这证明了 $I_B(t)=P(T^{-1}(B)|t)$, $a.s.P^T$. 另一方面,根据条件概率分布的定义, 有 $P(T^{-1}(B), t)=P(T^{-1}(B)|t)$, $a.s.P^T$, 于是由条件概率的 $a.s.P^T$ 的唯一性,即得(5.23).

现以 B_1, B_2, $\cdots$ 记 R^1 的有理区间全体, 对每个 i, 存在 P^T-零集 N_i,使当 $t\overline{\in} N_i(i=1, 2, \cdots)$时,有

$$P(T^{-1}(B_i), t)=I_{B_i}(t).$$

记 $N=\cup_i N_i$,则 N 为 P^T-零集, 且当 $t\overline{\in} N$ 时, (5.23)式当 B 为任意 B_i 时成立. 由于(5.23)式两边作为 $B\in\mathscr{B}_T$ 的函数, 都是 $\mathscr{B}_T$ 的概率测度,因此, 对 $t\overline{\in} N$ 及任意 $B\in\mathscr{B}_T$, 有(5.23). 在(5.23)中取 $t\overline{\in} N$, 而 $B=\{t\}$,即得(5.22). 定理证毕.

这个定理给"条件化"以一个很直观的意义: 给定条件 "$T(X)=t$",相当于把原概率空间$(\mathscr{X}, \mathscr{B}_x, P)$转化为概率空间$(A_t, \mathscr{B}_{xt}, P(\cdot,t))$, 这里 $\mathscr{B}_{xt}$ 是由 A_t 的子集构成的 σ-域$\{A\cap A_t: A\in\mathscr{B}_x\}$. 在下面的特殊情况下,这个结果有很醒目的形式.

设变量 Y、T 的样本空间分别为欧氏样本空间 $(\mathscr{Y}, \mathscr{B}_y)$ 和 $(\mathscr{T}, \mathscr{B}_{\mathscr{T}})$, $X=(Y, T)$, X 的样本空间为 $(\mathscr{X}, \mathscr{B}_x)=(\mathscr{Y}\times\mathscr{T}, \mathscr{B}_y\times\mathscr{B}_{\mathscr{T}})$. 这时有 $A_t=\mathscr{Y}\times\{t\}$. 因此,在 t 已给定时,不妨把 A_t 看成是 $\mathscr{Y}$. 给定 $T=t$ 时 X 的条件概率分布, 也就是给定 $T=t$ 时, Y 的条件概率分布 $P^{Y|t}$ 定义为

$$P^{Y|t}(C)=P(C\times\{t\}, t),$$

这里 $P(A, t)$就是给定 T 之下 X 的条件概率分布函数.

现设 $f(x)=f(y, t)$满足条件(5.1),则

$$E[f(X)]=E[f(Y, T)]=E\{E[f(Y, T)|T]\}. \tag{5.24}$$

而依定理 5.2, 有

$$E[f(Y, T)|t]=\int_{\mathscr{Y}} f(y, t)P^{Y|t}(dy), \quad a.s.P^T.$$

由此及(5.24),得

$$E[f(X)]=\int_{\mathscr{T}}\int_{\mathscr{Y}} f(y, t)P^{Y|t}(dy)dP^T(t). \tag{5.25}$$

这就是一般情况下(非独立情况下)的 Fubini 定理.

最后, 我们提出如下的注记: 上面在考虑条件概率分布时, 我

们指的是 X 本身的条件概率分布. 当然也可以考虑任一随机变量 $f(X)$ 的条件概率分布, $f(x)$ 是 $(\mathscr{X}, \mathscr{B}_x)$ 到 $(\mathscr{F}, \mathscr{B}_F)$ 的可测变换:"在给定 T 时 $f(X)$ 的条件概率分布"定义为 $f^{-1}(\mathscr{B}_F)\times\mathscr{T}$ 上的函数 $P(A, t)$, 满足如下条件:

1° 对任意的 $A\in f^{-1}(\mathscr{B}_F)$, 有

$$P(A, t)=P(A|t),\ a.s.P^T;$$

2° 当给定 t 时, $P(A, t)$ 作为 A 的函数, 是 $f^{-1}(\mathscr{B}_F)$ 上的概率测度.

§6 充分统计量

一、基本概念

1. 充分统计量的定义 在§4中我们引进过统计量的概念. 统计量是对样本的整理和加工. 经过整理、加工后, 把一大堆原始数据——样本 x 变换为比较简单的统计量 $T(x)$. 因此自然有可能在整理、加工过程中失掉一些有关未知参数 θ 的信息. 而这样带来的后果是: 如果我们基于统计量 $T(x)$ 对 θ 进行推断, 其效能可能达不到基于原始样本 x 对 θ 所作的推断.

举一很简单的例子: 设 $X_1, \cdots, X$ 为抽自正态总体 $N(\theta,1)$ 的 iid. 样本, 要估计 θ. 考虑统计量

$$T_m(x)=(x_1, \cdots, x_m),\ 1<m<n,$$

这等于只保留原始样本中最初 m 个观察值, 因而就丢掉了后 $n-m$ 个观察结果中所包含的关于 θ 的信息. m 愈小, 信息损失愈严重.

因此, 我们希望, 在将原始样本 x 整理为统计量 $T(x)$ 时, 信息损失愈少愈好. 如果统计量 $T(x)$ 保留了原始样本中所包含的关于未知参数 θ 的全部信息, 就称它是*充分统计量*.

为了给充分统计量下一个严格的数学定义, 必须找到一种数学工具, 以表达出"信息量无丢失"这个模糊的概念. 我们可以这样来分析问题: 原始样本 x 中所包含的(关于参数 θ 的, 下同)信息,

由两部分构成，一部分是统计量 $T(x)$ 所包含的信息，另一部分是在已知 $T(x)=t$ 条件下，由于进一步知道 x（设想：在将 x 加工为 $T(x)$ 后，就把 x 丢掉了．形象地，可以把情况设想成：观察者甲知道 x，并由甲将 x 加工为 $T(x)$，但乙只从甲那里知道 $T(x)$ 而不知道 x 带来的附加信息．这后一部分就是在将 x 加工为 $T(x)$ 时所损失掉的信息．因此，要 T 为充分统计量，则必须使后面这部分信息量为零．

但是，怎样去判断这部分信息量是否为零呢？考察在给定 T 的条件下，X 的条件概率分布．因为 X 的无条件分布为 P_θ，这个条件概率分布一般说来与 θ 有关．若情况果真是如此，则在已知 $T(x)$ 的条件下进一步得到 x，相当于从一族与 θ 有关的分布中获得样本，而这当然会包含与 θ 有关的信息．反过来，若上述条件概率分布与 θ 无关，则在已知 $T(x)$ 的条件下进一步得到 x，不过是从一与 θ 无关的分布中去抽样，它当然不会提供关于 θ 的任何信息．

这样一来，我们就抓住了充分统计量这个重要概念的实质所在，且导致下面的正式定义：

定义 6.1 设变量 X 的概率空间为 $(\mathscr{X}, \mathscr{B}_x, P_\theta, \theta\in\Theta)$，$T$ 为定义在 $(\mathscr{X}, \mathscr{B}_x)$ 上，值域为 $(\mathscr{T}, \mathscr{B}_T)$ 的统计量．若存在定义于 $\mathscr{B}_x\times\mathscr{T}$ 上的与 θ 无关的函数 $g(A, t)$，使对任何 $\theta\in\Theta$ 和 $A\in\mathscr{B}_x$，有

$$P_\theta(A|t)=g(A, t), \quad a.s. P_\theta^T. \tag{6.1}$$

P_θ^T 是 P_θ 经过统计量 T 导出的测度，$P_\theta(A|t)$ 是在 X 的分布为 P_θ 时，在给定 T 的条件下的条件概率，则称 T 为充分统计量．

(6.1)式的意义就在于此条件概率与 θ 无关．

我们来考察几个例子：

例 6.1 考察例 5.1，在此例中，$X=(X_1, \cdots, X_n)$，$X_1, \cdots, X_n$ 为 *iid.*，X_1 的分布为 F，F 属于某个一维分布族 $\mathscr{F}$．T 为次序统计量．由(5.9)知，给定 T 时 X 的条件概率与 F 无关（在(5.9)式中取 $f(x)=I_A(x)$，$A\in\mathscr{B}_x$）．因此，在本例中

序统计量是充分统计量．这个理论结果相应于如下的直观事实：在各次观察都在同样条件下独立进行时，观察结果的次序没有什么作用，即：我们只需知道全部 n 次观察结果，而无需确切知道哪个结果是在第几次观察中得到的．

例 6.2 设 $X=(X_1, \cdots, X_n)$, $X_1, \cdots, X_n$ 为 *iid.*，且 X_1 有两点分布

$$P_\theta(X_1=x_1)=\theta^{x_1}(1-\theta)^{1-x_1}, \quad (x_1=0, 1, 0\leqslant\theta\leqslant1). \tag{6.2}$$

这相当于从一批废品率为 θ 的产品中有放回地抽出 n 个，以估计 θ．从直观上显然，如果我们知道了 $T(x)=\sum\limits_1^n x_i$，即样品中总的废品个数，则进一步知道究竟这些废品是在哪几次抽得的，已经没有意义．也就是说，$T(x)=\sum\limits_1^n x_i$ 是一充分统计量．这个事实不难由定义 6.1 严格证明．事实上，有

$$P_\theta(X_1=x_1, \cdots, X_n=x_n|T=t)=\binom{n}{t}^{-1}, \tag{6.3}$$

这里 $t=0, 1, \cdots, n$, $x_i=0$ 或 1，且 $\sum\limits_1^n x_i=t$．在其它情况，上述条件概率为 0．因为 (6.3) 式的右边与 θ 无关，这证明了 T 的充分性．

例 6.3 设 $X=(X_1, \cdots, X_n)$, $X_1, \cdots, X_n$ 为 *iid.*，且 $X_1\sim N(\theta, 1)$．为估计 θ，一般用 $\overline{X}=\frac{1}{n}\sum\limits_1^n X_i$．我们在直观上感到：在知道了 $\overline{X}$ 后，进一步知道原始样本 X 已没有什么作用，虽然这个直观并非明显．下面我们证明，$T(x)=\bar{x}$ 确为充分统计量．

如果我们直接去计算在给定 T 时 X 的条件概率分布，则计算比较复杂．我们用如下的办法：作正交变换

$$Y_1=\sqrt{n}\,\overline{X},$$

$$Y_i=\sum_{j=1}^n c_{ij}X_j \quad (i=2, \cdots, n).$$

率论中周知的事实，$Y_1, \cdots, Y_n$ 独立，且 $Y_i\sim N(0, 1)$

$(i=2, \cdots, n)$. 显然，$\bar{X}$ 对原样本 X 的充分性等价于 Y_1 对$(Y_1, \cdots, Y_n)$ 的充分性，而这立即从定义得出．因为 $Y_1, \cdots, Y_n$ 是独立的，故给定 Y_1 时 $(Y_2, \cdots, Y_n)$ 的条件概率分布就是 $(Y_2, \cdots, Y_n)$的无条件分布，而后者为 $n-1$ 个标准正态分布的卷积，与 θ 完全无关.

在结束这一段以前，我们给出关于统计量的充分性的几点注记．设变量 X 的概率空间为$(\mathscr{X}, \mathscr{B}_x, P_\theta, \theta\subset\Theta)$.

1° 若统计量

$$T:(\mathscr{X}, \mathscr{B}_x)\mapsto(\mathscr{T}, \mathscr{B}_T)$$

有充分性，而 $\Theta'\subset\Theta$，则将 X 的分布族压缩为$(P_\theta, \theta\in\Theta')$后，$T$ 仍有充分性.

证明是显然的.

2° 设统计量 T 充分，而统计量 $S:(\mathscr{X}, \mathscr{B}_x)\to(\mathscr{S}, \mathscr{B}_s)$ 与 T 等价(统计量等价性的定义见(4.4)式，及该式上面一段的说明)，则 S 也是充分统计量.

证 由统计量 S 与 T 等价，直接根据条件概率的定义，即得 $P_\theta(A|S)=P_\theta(A|T)|_{T=\psi(S)}$. 于是由 $P_\theta(A|T)$ 与 θ 无关即得 $P_\theta(A|S)$与 θ 无关.

这个性质基本上是说：对充分统计量进行一一对应变换不丧失充分性．如在例 6.3 中，由 $\bar{X}$ 的充分性可推出 $e^{\bar{x}}$ 与 $\left(\sum_1^n X_i\right)^3$ 都是充分统计量，等等.

2. 因子判别法　直接从定义去验证一个统计量的充分性是很麻烦的，因为涉及条件概率的计算．不过，好在关于统计量的充分性有一个很一般的判别法．这个判别法的比较简单的情况由 H. Cramer 在 [3] 中证明了，一般情况是 Halmos 和 Savage 在 1949 年证明的(见[2])．本段我们先提出这个判别法并举例说明其应用．严格的证明放在 6.3 段．初学者在第一遍读时可略过这个证明，不影响对本书以后的理解.

定理 6.1(因子判别定理)　设 T 为由$(\mathscr{X}, \mathscr{B}_x, P_\theta, \theta\in\Theta)$到

$(\mathscr{T}, \mathscr{B}_T)$的统计量，又设存在 $\mathscr{B}_x$ 上的 σ-有限测度 μ，使 $P_\theta \ll \mu$ 对一切 $\theta \in \Theta$. 以 $P_\theta(x)$ 记 P_θ 对 μ 的 Radsm-Nikodym 导数 $dP_\theta(x)/d\mu$. 则 T 为充分统计量的充要条件是：存在定义于 $\mathscr{X}$ 上的非负 $\mathscr{B}_x$-可测函数 $h(x)$，以及对任何 $\theta \in \Theta$，存在定义于 $\mathscr{T}$ 上的 $\mathscr{B}_T$-可测函数 $g_\theta(t)$，使得 $\theta \in \Theta$ 有

$$p_\theta(x) = g_\theta(T(x))h(x), \quad a.s.\mu. \tag{6.4}$$

注意：上式中的例外集可与 θ 有关.

例 6.4 设 X 服从本章(2.1)式所定义的指数族分布. 记

$$T(x) = (T_1(x), \cdots, T_k(x)), \tag{6.5}$$

则由因子判别定理可得出 T 是充分统计量. 不难看出，例 6.2 与 6.3 都是此例的特殊情况. 又如 $X_1, \cdots, X_n$ 为 iid., $X_1 \sim N(a, \sigma)$, $\theta = (a, \sigma)$，则 $T = (\overline{X}, S^2)$ 为充分统计量，其中 $S^2 = \sum_1^n (x_i - \overline{X})^2$. 又在例 2.2 中，若 β 已知，而 $\theta = \alpha$，则 $T(x) = \sum_1^n x_i$ 为充分统计量. 若 α、β 都未知即 $\theta = (\alpha, \beta)$，则 $T(x) = \left(\sum_1^n x_i, \sum_1^n \log x_i\right)$ 为充分统计量，等等. 由于很多常见的分布族都是指数族，故本例有很大的意义.

由因子判别定理得出指数族一个重要特性：不论样本大小 n 多么大，指数族有一固定维数的充分统计量. 实际上，若 $X_1, \cdots, X_n$ 为取自分布族(2.1)的 *iid.* 样本，$X = (X_1, \cdots, X_n)$，则 X 的分布为

$$p_\theta^*(x)d\mu^{(n)} = C^n(\theta)\exp\left[\sum_1^k Q_j(\theta)T_j^{(n)}(x)\right]d\mu^{(n)},$$

此处 $\mu^{(n)} = \mu \times \cdots \times \mu$, $T_j^{(n)}(x) = \sum_{i=1}^n T_j(x_i)$. 由因子判别定理得出：$T^{(n)}(x) = (T_1^{(n)}(x), \cdots, T_k^{(n)}(x))$ 为充分统计量，其维数 k 与 n 无关. 在一些补充假定下可以证明：这个性质是指数族的特征性质.

下面举几个例子，其分布族中的分布不具备共同的负荷集.

例 6.5 设 $X_1, \cdots, X_n$ 为抽自均匀分布 $R(0, \theta)$ 的 *iid.* 样

本，$\theta>0$．$X=(X_1, \cdots, X_n)$的密度函数(对 L 测度)为

$$p_\theta(x)=\theta^{-n}I_{\{0<\max\limits_{1\leqslant i\leqslant n} x_i<\theta\}}(x).$$

由此表示式用因子判别定理，即知 $T(x)=\max\limits_{1\leqslant i\leqslant n} x_i$ 为一充分统计量．

例 6.6 设 $X_1, \cdots, X_n$ 为抽自负指数分布

$$f(x;\alpha,\beta)=\begin{cases}\beta^{-1}\exp\left(-\dfrac{x-\alpha}{\beta}\right), & x>\alpha;\\ 0, & x\leqslant\alpha\end{cases}$$

的 iid. 样本．$\beta>0$，$-\infty<\alpha<\infty$，$\theta=(\alpha,\beta)$．这个分布在讨论元件的寿命问题时很常用．在这种解释下，可以看成是有 n 个元件参加寿命试验，第 i 个元件的寿命为 X_i．

在所谓"截尾试验"中，我们只观察到一定个数(如 r 个)元件失效为止，也就是说，若以$(X_{(1)}, \cdots, X_{(n)})$记 $X_1, \cdots, X_n$ 的次序统计量，则我们实际只观察了 $X_{(1)}, \cdots, X_{(r)}$．不难证明(具体验证留给读者作为练习)$(X_{(1)}, \cdots, X_{(r)})$的密度函数(对 L 测度)为

$$f(x_{(1)},\cdots,x_{(r)};\theta)=\begin{cases}\dfrac{n!}{(n-r)!}\beta^{-r}\exp\left\{-\dfrac{\sum\limits_1^r x_{(i)}+(n-r)x_{(r)}-n\alpha}{\beta}\right\}, & \text{当 } x_{(1)}>\alpha;\\ 0, & \text{当 } x_{(1)}\leqslant\alpha\end{cases}$$

$$=\frac{n!}{(n-r)!}\beta^{-r}e^{-n\alpha/\beta}\exp\left\{-\frac{\sum\limits_1^r x_{(i)}+(n-r)x_{(r)}}{\beta}\right\}\cdot I_{(\alpha,\infty)}(x_{(1)}).$$

由此表达式并用因子判别定理，即得统计量$\left(X_{(1)}, \sum\limits_1^r X_{(i)}+(n-r)X_{(r)}\right)$充分．因而与之等价的统计量$\left(X_{(1)}, \sum\limits_2^r X_{(i)}+(n-r)X_{(r)}-(n-1)X_{(1)}\right)$也充分．

二、充分性原则

1．引言和一个引理 前面讲述统计量的充分性时，我们是从

“不丧失样本中包含的关于 θ 的信息量”这一角度着眼的. 但是,将样本简化的统计量,不是为简化而简化,而是为了将其用于统计判决问题. 既然充分统计量包含了原始样本的全部信息,因而自然地使我们有理由认为:使用充分统计量代替原样本以进行统计推断不应使我们丧失任何东西. 更确定地说:凡是用原始样本能达到的,用充分统计量也能达到. 所有关于这一点的正面论断都称之为**充分性原则**.

在确切陈述并证明充分性原则以前,我们需要一个引理.

引理 6.1 沿用前面的记号. 设样本空间 $(\mathscr{X}, \mathscr{B}_x)$ 为欧氏的,而 T 为充分统计量,则存在定义于 $\mathscr{B}_x\times\mathscr{T}$ 上的函数 $P(A, t)$,满足以下条件:

1° 对任何固定的 $t\in\mathscr{T}$, $P(A, t)$ 作为 $\mathscr{B}_x$ 上的函数,是一个概率测度.

2° 对任何 $\theta\in\Theta$ 及 $A\in\mathscr{B}_x$,有

$$P_\theta(A|t)=P(A, t), \quad a.s. P_\theta^T.$$

根据前面关于条件概率分布的定义,这个引理无非是说:在样本空间为欧氏的时候,对一切 θ 存在公共的(在给定 T 时)条件概率分布.

本引理的证明与定理 5.1 的证明完全相似. 只需注意,由于 T 的充分性,定理 5.1 证明中的 $P((-\infty, x]|t)$ 在此处用 $P_\theta((-\infty, x]|t)$ 代替时,其结果与 θ 无关,故仍可记为 $F(x, t)$. 证明的其余部分无差别.

2. 随机化判决函数下的充分性原则

定理 6.2 设样本空间 $(\mathscr{X}, \mathscr{B}_x)$ 为欧氏的,T 是一充分统计量,$\mathscr{D}$ 是[以 $(\mathscr{X}, \mathscr{B}_x, P_\theta, \theta\in\Theta)$ 为概率空间的]某一统计判决问题的全部(包括随机化的,下同)判决函数,而 $\mathscr{D}_T$ 为只依赖于 T 的全部判决函数,则 $\mathscr{D}_T$ 构成 $\mathscr{D}$ 的一个本质完全类.

这个定理确切地表述了“充分性原则”的含义:当统计量 T 为充分时,若使用某个基于原始样本的判决函数 δ 得到风险函数 $R(\theta, \delta)$,则可以找到一个只依赖于 T 的判决函数 δ_T,其风险函数

$R(\theta, \delta_T)$ 满足条件 $R(\theta, \delta_T) \leqslant R(\theta, \delta)$ 对一切 $\theta \in \Theta$. 这说明: 在将原样本简化为统计量 T 时, 从风险的角度看不致丧失任何信息.

定理的证明 以 $(A, \mathscr{B}_A)$ 记行动空间, 用 $L(\theta, a)$ 记损失函数. 任取判决函数 $\delta(D|x)$. 根据(1.2), 其风险函数为

$$R(\theta, \delta) = \int_{\mathscr{X}} \int_A L(\theta, a) \delta(da|x) dP_\theta(x).$$

由于 T 为充分统计量, 且 $(\mathscr{X}, \mathscr{B}_x)$ 为欧氏的, 根据引理 6.1, 存在公共的(对一切 θ 有效的)条件概率分布 $P(A, t)$. 引进只依赖于 T 的判决函数 δ^* 如下:

$$\delta^*(D|t) = \int_{\mathscr{X}} \delta(D|x) P(dx, t).$$

关于 δ^* 确为判决函数的事实, 不难根据 1.2 节(一)中关于判决函数的可测性的要求来证明(建议初学者仔细写出这个验证过程). 有

$$\begin{aligned} R(\theta, \delta^*) &= \int_{\mathscr{T}} \int_A L(\theta, a) \delta^*(da|t) dP_\theta^T(t) \\ &= \int_{\mathscr{T}} \int_A L(\theta, a) \left[\int_{\mathscr{X}} \delta(da|x) P(dx, t) \right] dP_\theta^T(t) \\ &= \int_{\mathscr{T}} \int_{\mathscr{X}} \int_A L(\theta, a) \delta(da|x) P(dx, t) dP_\theta^T(t) \\ &= \int_{\mathscr{X}} \int_A L(\theta, a) \delta(da|x) dP_\theta(x) = R(\theta, \delta). \end{aligned}$$

在这里, 第三步用了一般的 Fubini 定理((5.25)式), 第四步用了定理 4.2. 定理证毕.

这个定理的好处是: 它对损失函数和行动空间没有任何要求.

3. 非随机化判决函数下的充分性原则

定理 6.3 设样本空间 $(\mathscr{X}, \mathscr{B}_x)$ 为欧氏的, 而行动空间 $(A, \mathscr{B}_A) = (R^m, \mathscr{B}^m)$, 对某个自然数 m. 又设损失函数 $L(\theta, a)$ 有如下的性质:

1° 对任何 $\theta \in \Theta$, $L(\theta, a)$ 为 R^m 上的凸函数(在(3.5)式的意义下).

2° 对任何 $\theta \in \Theta$, 有 $\lim\limits_{\|a\| \to \infty} L(\theta, a) = \infty$.

设 T 为一充分统计量，$\mathscr{D}$ 为一切非随机化的判决函数的类，而 $\mathscr{D}_T$ 为一切只依赖于 T 的非随机化的判决函数的类，则 $\mathscr{D}_T$ 构成 $\mathscr{D}$ 的本质完全类.

证 根据引理 6.1，存在与 θ 无关的条件概率分布 $P(A, t)$. 任取 $\delta \in \mathscr{D}$. 令

$$C = \left\{ t: \int_{\mathscr{X}} \| \delta(x) \| P(dx, t) < \infty \right\},$$

然后定义只依赖于 t 的非随机化的判决函数

$$\delta^*(x) = \begin{cases} \int_{\mathscr{X}} \delta(x) P(dx, t), & \text{当 } t \in C; \\ 0, & \text{当 } t \overline{\in} C, \end{cases} \quad T(x) = t,$$

则由 Jensen 不等式(定理 3.7)有

$$L(\theta, \delta^*(t)) \leqslant \int_{\mathscr{X}} L(\theta, \delta(x)) P(dx, t), \text{当 } t \in C. \quad (6.6)$$

又由本定理假设，用定理 3.5，可知对任何 $\theta \in \Theta$，存在常数 $h_\theta > 0$ 及 b_θ，使 $L(\theta, a) \geqslant h_\theta \|a\| + b_\theta$. 于是当 $t \overline{\in} C$ 时，有

$$\int_{\mathscr{X}} L(\theta, \delta(x)) P(dx, t) \geqslant h_\theta \int_{\mathscr{X}} \| \delta(x) \| P(dx, t) + b_\theta = \infty. \quad (6.7)$$

由(6.6)及(6.7)知，对一切 $t \in \mathscr{T}$，有

$$L(\theta, \delta^*(t)) \leqslant \int_{\mathscr{X}} L(\theta, \delta(x)) P(dx, t),$$

因而

$$\begin{aligned} R(\theta, \delta^*) &= \int_{\mathscr{T}} L(\theta, \delta^*(t)) dP_\theta^T(t) \\ &\leqslant \int_{\mathscr{T}} \int_{\mathscr{X}} L(\theta, \delta(x)) P(dx, t) dP_\theta^T(t) \\ &= \int_{\mathscr{X}} L(\theta, \delta(x)) dP_\theta(x) = R(\theta, \delta). \end{aligned}$$

定理得证.

可以举例证明(见第四章例)：当定理中的条件不成立时，本定理的结论可以不对. 现在还不完全清楚的是：在非随机化判决函数类中的充分性原则还在哪些其他情况下有效.

可以证明：在一定条件下，充分性原则的逆亦成立，即：若一统计量 T 具有在充分性原则中所表述的性质，则 T 必是充分的．在此我们不深入讨论这个问题的细节了．

三、* 因子判别定理的证明

这个定理的证明比较长．先作一些准备工作．设 $\{\mu\}$ 和 $\{\nu\}$ 为 $(\mathscr{X}, \mathscr{B}_x)$ 上的两个测度族（μ, ν 分别是这两个族中的"代表"）．如果由条件：对任一 $A\in\mathscr{B}_x$,

$$\text{对每一 } \mu\in\{\mu\},\ \text{有 } \mu(A)=0.$$

可推出 $$\text{对每一 } \nu\in\{\nu\},\ \text{有 } \nu(A)=0.$$

则称测度族 $\{\nu\}$ 为测度族 $\{\mu\}$ 所控，记为 $\{\nu\}\ll\{\mu\}$．若 $\{\nu\}\ll\{\mu\}$ 及 $\{\mu\}\ll\{\nu\}$，则称 $\{\mu\}$ 与 $\{\nu\}$ 等价，记为 $\{\mu\}\sim\{\nu\}$．

容易证明，若 μ 为 $\mathscr{B}_x$ 上的 σ-有限测度，且 $\{\nu\}\ll\mu$（这里 μ 表示只包含一个测度 μ 的族），则必存在 $\mathscr{B}_x$ 上的概率测度 μ^*，使 $\{\nu\}\ll\mu^*$．事实上，不失普遍性，可设 $\mu(\mathscr{X})>0$（否则，结论显然成立），这时，由于 μ 为 σ-有限，存在一串两两不相交的 $\{A_i\}$，$\bigcup_{i=1}^{\infty} A_i=\mathscr{X}$，使 $0<\mu(A_i)<\infty$ $(i=1, 2, \cdots)$．取

$$\mu^*(A)=\sum_{1}^{\infty}\frac{\mu(A\cap A_i)}{2^i\mu(A_i)}\quad (A\in\mathscr{B}_x), \tag{6.8}$$

则易见 μ^* 即为具有上述性质的概率测度．

引理 6.2 设 $\mathscr{P}=(P_\theta,\ \theta\in\Theta)$ 为 $\mathscr{B}_x$ 上的一族概率测度，则 $\mathscr{P}$ 为 $\mathscr{B}_x$ 上某一 σ-有限测度 μ 所控的充要条件为：存在 Θ 的可列子集 Θ'，使

$$\mathscr{P}\sim(P_\theta,\ \theta\in\Theta') \tag{6.9}$$

证 充分性显然，因若(6.9)成立，则将 Θ' 中的元排列为 θ_1, θ_2, $\cdots$，取 $\mu=\sum_{i=1}^{\infty}P_{\theta_i}/2^i$ 即可．

现证必要性：设 $\mathscr{P}\ll\mu$，μ 为 σ-有限测度．由上面的说明，可设 μ 为概率测度．引进测度族

$$\mathscr{Q}=\{Q\colon Q=\sum_i a_iP_{\theta_i},\ \theta_i\in\Theta,\ a_i>0,\ \sum_i a_i=1\},$$

易见 $\mathscr{P}\sim\mathscr{Q}$．对每个 $Q\in\mathscr{Q}$，记 $q(x)=dQ(x)/d\mu$，再令

$$\mathscr{F}=\{C\colon C\in\mathscr{B}_x, \text{存在 } Q\in\mathscr{Q},\ \text{使 } \mu(C\cap\{x\colon q(x)=0\})=0\}, \tag{6.10}$$

在 $\mathscr{F}$ 中取出序列 $\{C_i\}$，使

$$\lim_{i\to\infty}\mu(C_i)=\sup\{\mu(C)\colon C\in\mathscr{F}\}<\infty. \tag{6.11}$$

在 $\mathscr{F}$ 的定义中，对每个 $C\in\mathscr{F}$，都有某个 $Q\in\mathscr{Q}$ 与之相应．设与 C_i 相应的测

度为 Q_i, 而 $q_i=dQ_i/d\mu$. 再令

$$C_0=\bigcup_1^\infty C_i,\ Q_0=\sum_1^\infty Q_i/2^i,\ q_0=\sum_1^\infty q_i/2^i,$$

则 $Q_0\in\mathscr{Q}$, 且 $q_0=dQ_0/d\mu$, 且

$$\mu(C_0)\geqslant\sup\{\mu(C):C\in\mathscr{F}\},\tag{6.12}$$

又因

$$\mu(C_0\cap\{x:q_0(x)=0\})\leqslant\sum_1^\infty\mu(C_i\cap\{x:q_i(x)=0\})=0,\tag{6.13}$$

可知 $C_0\in\mathscr{F}$. 因此, (6.12)式实际上等号成立.

显然, 若能证明 $\mathscr{Q}\ll Q_0$, 也就证明了 $(P_\theta,\ \theta\in\Theta)\ll\{P_{\theta_i},\ i=1,\ 2,\ \cdots\}$. 设 $A\in\mathscr{B}_x$ 且 $Q_0(A)=0$. 对任一取定的 $Q\in\mathscr{Q}$, 记

$$C=\{x:q(x)>0\},\tag{6.14}$$

则因 $Q_0(A\cap C_0)=0$, 有

$$0=Q_0(A\cap C_0\cap\{x:q_0(x)>0\})=\int_{A\cap C_0\cap\{x:q_0(x)>0\}}q_0(x)d\mu,$$

可知 $$\mu(A\cap C_0\cap\{x:q_0(x)>0\})=0.$$

将此式与(6.13)结合起来, 得 $\mu(A\cap C_0)=0$. 再由 $\mathscr{Q}\ll\mu$, 有

$$Q(A\cap C_0)=0.\tag{6.15}$$

再由 C 的定义(6.14), 得

$$Q(A\cap\overline{C}_0\cap\overline{C})=0\tag{6.16}$$

($\overline{C}_0=\mathscr{X}-C_0$, 等等). 若能证明

$$Q(A\cap\overline{C}_0\cap C)=0,\tag{6.17}$$

则由(6.15)～(6.17)得 $Q(A)=0$. 这将得出 $Q\ll Q_0$. 因为 Q 是从 $\mathscr{Q}$ 中任取的, 得 $\mathscr{Q}\ll Q_0$. 如上所指出的, 这将完成引理的证明.

用反证法: 设(6.17)式不成立, 即 $Q(A\cap\overline{C}_0\cap C)>0$. 记

$$Q^*=\frac{1}{2}(Q_0+Q),\quad q^*=\frac{1}{2}(q_0+q),\ C^*=C_0\cup(A\cap\overline{C}_0\cap C).$$

显然, $Q^*\in\mathscr{Q}$. 又因 $q^*=dQ^*/d\mu$, 及

$$\mu(C^*\cap\{x:q^*(x)=0\})\leqslant\mu(C_0\cap\{x:q_0(x)=0\})=0\tag{6.18}$$

[(6.18) 的后一式根据 (6.13); 前一式根据 $\{x:q^*(x)=0\}\subset\{x:q_0(x)=0\}$, $\{x:q^*(x)=0\}\subset\{x:q(x)=0\}$ 以及 $C\cap\{x:q(x)=0\}=\{x:q(x)>0\}\cap\{x:q(x)=0\}=\phi$] 可知 $C^*\in\mathscr{F}$.. 于是由 (6.12) 及 $Q(A\cap\overline{C}_0\cap C)>0$ [它推出 $\mu(A\cap\overline{C}_0\cap C)>0$]有

$$\mu(C^*)=\mu(C_0)+\mu(A\cap\overline{C}_0\cap C)>\mu(C_0)\geqslant\sup\{\mu(C):C\in\mathscr{F}\}.$$

这与 $C^*\in\mathscr{F}$ 矛盾. 从而证明了(6.17), 于是完成了引理的证明.

引理 6.3 设 $(\mathscr{X},\ \mathscr{B}_x)$ 上的分布族 $(P_\theta,\ \theta\in\Theta)$ 等价于其一可列子族 $\{P_{\theta_i},$

$i=1, 2, \cdots\}$. 设 $\lambda=\sum_1^\infty a_i P_{\theta_i}$, 此处 $a_i>0$ 而 $\sum_1^\infty a_i=1$. 设 T 为定义于 $\mathscr{X}$ 上而以$(\mathscr{T}, \mathscr{B}_T)$为值域的统计量. 则 T 为充分统计量的充要条件是: 对每个 $\theta\in\Theta$,存在 $\mathscr{B}_T$-可测的非负函数 $g_\theta(t)$,使

$$dP_\theta(x)/d\lambda=g_\theta(T(x)),\ a.s.\lambda. \tag{6.19}$$

证 记 $\mathscr{B}_0=T^{-1}(\mathscr{B}_T)$. 为证充分性, 只需证明对任何 $\theta\in\Theta$ 及 $A\in\mathscr{B}_x$, 有

$$\lambda(A|T(x))=P_\theta(A|T(x)),\ a.s.P_\theta. \tag{6.20}$$

此式证明如下: 任取 $B\in\mathscr{B}_0$,有

$$\begin{aligned}P_\theta(A\cap B)&=\int_{A\cap B}dP_\theta(x)=\int_{A\cap B}g_\theta(T(x))d\lambda\\&=\int_B I_A(x)g_\theta(T(x))d\lambda\\&=\int_B E_\lambda[I_A(x)g_\theta(T(x))|T(x)]d\lambda\\&=\int_B E_\lambda[I_A(x)|T(x)]g_\theta(T(x))d\lambda\\&=\int_B E_\lambda[I_A(x)|T(x)]dP_\theta.\end{aligned}$$

这里第二步用 (6.19), 第四步根据条件期望的定义及 $B\in\mathscr{B}_0$, 最后一步是因为: $g_\theta(T(x))$为 $\mathscr{B}_0$-可测,故由(6.19)知: 在基本概率空间为$(\mathscr{X}, \mathscr{B}_0, P_\theta)$时,仍有 $dP_\theta(x)/d\lambda=g_\theta(T(x))$, $a.s.\lambda$.

现证必要性: 设 T 为 $(\mathscr{X}, \mathscr{B}_x, P_\theta, \theta\in\Theta)$ 之下的充分统计量. 任取 $A\in\mathscr{B}_x$, 存在与 θ 无关的条件概率 $P(A|t)$. 对任意的 $B\in\mathscr{B}_0$,有

$$\int_B P(A|T(x))dP_\theta(x)=P_\theta(A\cap B).$$

由于此式对任何 $\theta\in\Theta$ 成立,考虑到 λ 的定义,有

$$\int_B P(A|T(x))d\lambda(x)=\lambda(A\cap B).$$

因此,在基本概率空间为$(\mathscr{X}, \mathscr{B}_x, \lambda)$时, $P(A|t)$也是在给定 T 之下的条件概率. 根据定理 4.1,存在 $\mathscr{B}_T$-可测函数 $g_\theta(t)$, 使当基本概率空间为$(\mathscr{X}, \mathscr{B}_0)$时(即把 P_θ、λ 都看作 $\mathscr{B}_0$ 上的概率测度),有

$$dP_\theta(x)/d\lambda=g_\theta(T(x)),\ a.s.(\mathscr{B}_0, \lambda). \tag{6.21}$$

现在证明: 即使在基本空间取为$(\mathscr{X}, \mathscr{B}_x)$,仍有

$$dP_\theta(x)/d\lambda=g_\theta(T(x)),\ a.s.(\mathscr{B}_x, \lambda). \tag{6.21}$$

证明了(6.21)也就是证明了(6.19), 从而将完成引理的证明. 为证(6.21), 任取 $A\in\mathscr{B}_x$,有

$$P_\theta(A)=\int_{\mathscr{X}} P(A|T(x))dP_\theta(x)=\int_{\mathscr{X}} E_\lambda[I_A(X)|T(x)]dP_\theta(x)$$

$$=\int_{\mathscr{X}} E_\lambda[I_A(X)|T(x)]g_\theta(T(x))d\lambda(x)$$

$$=\int_{\mathscr{X}} E_\lambda[g_\theta(T(X))I_A(X)|T(X)]d\lambda(x)$$

$$=\int_{\mathscr{X}} I_A(x)g_\theta(T(x))d\lambda(x)$$

$$=\int_A g_\theta(T(x))d\lambda(x).$$

这就证明了(6.21). 在上面这一系列等式中,第一个是根据条件概率的定义,第二个是根据(6.20),第三步利用(6.21)以及 $E_\lambda[I_A(X)|T(x)]$ 为 $\mathscr{B}_0$-可测,最后两步是显然的. 引理证毕.

现在转到定理 6.1 的证明. 在定理 6.1 中的假定下,根据引理 6.2, 存在 $(P_\theta,\ \theta\in\Theta)$ 的可列子族 $\{P_{\theta_i},\ i=1, 2, \cdots\}$, 等价于 $(P_\theta,\ \theta\in\Theta)$. 记 $\lambda=\sum_1^\infty P_{\theta_i}/2^i$. 先设 T 为充分统计量,由引理 6.3 知

$$dP_\theta(x)/d\lambda=g_\theta[T(x)],\ a.s.\lambda.$$

又因 $\lambda\ll\mu$, 有

$$dP_\theta(x)/d\mu=(dP_\theta(x)/d\lambda)(d\lambda(x)/d\mu)=g_\theta(T(x))h(x),\quad a.s.\mu.$$

这里 $h(x)=d\lambda(x)/d\mu$. 这就证明了定理的必要性部分.

为证充分性,设条件(6.4)成立. 则

$$d\lambda(x)=\sum_1^\infty \frac{1}{2^i}g_{\theta_i}(T(x))h(x)d\mu(x)=K(T(x))h(x)\mu(x),\ a.s.\mu.$$

此处 $K(u)=\sum_1^\infty g_{\theta_i}(u)/2^i$. 令

$$g_\theta^*(t)=\begin{cases} g_\theta(t)/k(t), & \text{当 } k(t)>0\\ 0, & \text{当 } k(t)=0\end{cases}$$

再因 $\lambda\ll\mu$, $P_\theta\ll\lambda$, 由 Radon-Nikodym 导数性质,

$$\frac{dP_\theta(x)}{d\lambda}=\left(\frac{dP_\theta(x)}{d\mu}\right)\Big/\left(\frac{d\lambda(x)}{d\mu}\right)=g_\theta^*[T(x)],\quad a.s.\lambda.$$

于是根据引理 6.3 得知: T 为充分统计量. 定理证毕.

问题与习题

1. 设 $X_1, \cdots, X_n$ 为 iid. 随机变量, $X_1\sim N(b,\ \sigma^2)$, $-\infty<b<\infty, \sigma^2>0$, 问题是要估计均值 b. 在平方损失 $L((b,\ \sigma^2),\ a)=(b-a)^2$ 下, 试证: (i)不

存在一致最优的估计；(ii) $\hat{\delta}(x)\equiv c$（常数）是容许估计；(iii) 当 $|c|>1$ 时，$\hat{\delta}(x)=c\bar{X}$ 不是容许估计.

2. 设 $\mathscr{X}=\{0, 1, 2\}$，考虑其上二项分布族：$P_\theta\{X=i\}=\binom{2}{i}\theta^i(1-\theta)^{2-i}$，$i\in\mathscr{X}, \theta\in\left\{\frac{1}{4}, \frac{3}{4}\right\}$. 问题是要估计 θ. 取损失函数为 $L(\theta, a)=\begin{cases}0, 当\ a=\theta\\ 1, 当\ a\neq\theta,\end{cases}$ $a\in A=\left\{\frac{1}{4}, \frac{3}{4}\right\}$. 试分别求在随机化估计类中和非随机化估计类中的 Minimax 估计.

3. 设 $X_1, \cdots, X_n$ 为 iid. 随机变量，$X_1\sim N(\theta, 1)$，$-\infty<\theta<\infty$. 在平方损失下，证明 $\delta(x)=\bar{x}$ 为 θ 的 Minimax 估计.

4. 设 $(\Theta, \mathscr{B}_\Theta)$ 上有概率分布 $d\xi(\theta)=\rho(\theta)d\theta$，并设真参数值为 θ 时，X 的概率密度为 $p_\theta(x)dx$. 则判决函数 $\delta(x)$ 为 Bayes 解（对某一给定判决问题）；如果它使积分值 $\int L(\theta, \delta(x))\pi(\theta|x)d\theta$ 达到最小（这里 $\pi(\theta|x)=\rho(\theta)p_\theta(x)\Big/\int\rho(\theta')P_{\theta'}(x)d\theta'$)，则它为给定 x 时未知参数 θ 的条件（后验）概率密度.

5. 设参数空间 $\Theta=\Theta_H\cup\Theta_k$，$\Theta_H\cap\Theta_k=\emptyset$. 检验零假设 $H_0: \theta\in\Theta_H$；对立假设 $H_1: \theta\in\Theta_k$. 行动空间 $A=\{a_0, a_1\}$，其中 a_0、a_1 分别表示接受和拒绝零假设. 令损失函数为

$$L(\theta, a)=\begin{cases}0, 当\ \theta\in\Theta_H,\ a=a_0\ 或\ \theta\in\Theta_k,\ a=a_1;\\ c, 当\ \theta\in\Theta_k,\ a=a_0;\\ d, 当\ \theta\in\Theta_H,\ a=a_1,\end{cases}$$

其中 c、d 为正常数. 利用第 4 题来证明：如下定义的判决函数为 Bayes 解：当 $cp\{\theta\in\Theta_k|x\}<dP\{\theta\in\Theta_H|x\}$ 时，令 $\delta(x)=a_1$；否则，令 $\delta(x)=a_0$.

6. 设 $X\sim N(\theta, 1)$，$-\infty<\theta<\infty$. 设损失函数为

$$L(\theta, a)=\begin{cases}0, & |\theta-a|\leqslant 1;\\ 1, & |\theta-a|>1,\end{cases}$$

$a\in A=(-\infty, \infty)$. 又设 θ 的先验分布为 $N(0, \tau^2)$，$\tau>0$ 为常数. 求 Bayes 解及其 Bayes 风险.

7. 假设 X_1, X_2, X_3 相互独立，$X_i\sim R(0, \theta_i)$，$0<\theta_i<\infty\ (i=1, 2, 3)$. 要由 (x_1, x_2, x_3) 来判决 $\theta_1, \theta_2, \theta_3$ 谁最大. 行动空间 $A=\{a_1, a_2, a_3\}$，其中 a_i 表示认为 θ_i 最大 $(i=1, 2, 3)$. 取损失函数 $L((\theta_1, \theta_2, \theta_3), a_i)=3-r_i$，这里 r_i 为 θ_i 在 $\theta_1, \theta_2, \theta_3$ 中的秩，即当 $\theta_i=\max(\theta_1, \theta_2, \theta_3)$ 时，$r_i=3$；$\theta_i=\min\{\theta_1, \theta_2, \theta_3\}$ 时，$r_i=1$；其它情况 $r_i=2$. 作判决函数如下：$\delta(x_1, x_2, x_3)=a_i$，当 $x_i=\max\{x_1, x_2, x_3\}\ (i=1, 2, 3)$. 计算 δ 的风险函数.

8. 假设损失函数 $L(\theta, a)$ 为 $\mathscr{B}_\Theta\times\mathscr{B}_A$ 可测函数；$\delta(x)$ 为 $\mathscr{B}_x$ 可测函数. 对任一 $B\in\mathscr{B}_x$, $P_\theta(B)$ 为 θ 的 $\mathscr{B}_\Theta$-可测函数，并假定对每一 θ,

$$\int_{\mathscr{X}} L(\theta, \delta(x))dP_\theta(x)<\infty,$$

证明：它必为 θ 的 $\mathscr{B}_\Theta$ 可测函数.

9. 把 § 2 中的各例题写成指数族的标准形式(2.2).

10. 设单参数一维指数族 $c(\theta)e^{\theta x}d\mu(x)$ 的自然参数空间为 $\Theta=(a, \infty)$, μ 的负荷集的上确界为 b. 则对任何 $x_0<b$, 有 $\lim\limits_{\theta\to\infty}c(\theta)e^{\theta x_0}=0$.

11. 举出这样一些单参数一维指数族 $c(\theta)e^{\theta x}d\mu(x)$, 使: (i)自然参数空间为 $[a, \infty)$; (ii)自然参数空间为 (a, ∞); (iii)自然参数空间为 (a, b); (iv)自然参数空间为 $[a, b]$; (v)自然参数空间为 $[a, b]$, 当 $\theta=b$ 时, X 的 $r(>0)$ 阶矩有限，但对任何 $\varepsilon>0$, $r+\varepsilon$ 阶矩不存在.

12. 设单参数指数族 $c(\theta)e^{\theta x}d\mu(x)$ 的自然参数空间 Θ 的上确界为 a, 则有: (i)若 $a=\infty$, 当 $\mu(0, \infty)>0$ 时，有 $\lim\limits_{\theta\to a}c(\theta)=0$; 当 $\mu(0, \infty)=0$、$\mu(0)>0$ 时，有 $\lim\limits_{\theta\to a}c(\theta)=\dfrac{1}{\mu(0)}$; 其他情况，有 $\lim\limits_{\theta\to a}c(\theta)=\infty$. (ii)若 $a<\infty$, 当 $a\in\Theta$ 时，则 $\lim\limits_{\theta\to a}c(\theta)=A$ 存在有限，且 $A=c(a)$; 当 $a\bar{\in}\Theta$ 时, $\lim\limits_{\theta\to a}c(\theta)=0$.

若 Θ 的下确界为 b, 写出与上面相应的结果.

13. 二维指数族 $c(\theta_1, \theta_2)e^{\theta_1 x+\theta_2 y}d\mu(x, y)$ 的边缘分布是否一定为一维指数族?

14. 设 $c(\theta)e^{\theta x}d\mu(x)$ 为一维指数族，则它为正态分布族 $\Leftrightarrow$ 其方差与 θ 无关.

15. 设 $c(\theta)e^{\theta x}d\mu(x)$ 为一维指数族，则它为 Poisson 分布族 $\Leftrightarrow$ 其均值与方差相等.

16. 设 $f(x)$ 为定义于 R^1 上的实函数，其为连续凸函数 $\Leftrightarrow$ 对任意的 $x<y<z$, 有

$$\frac{f(y)-f(x)}{y-x}\leqslant\frac{f(z)-f(x)}{z-x}\leqslant\frac{f(z)-f(y)}{z-y}.$$

17. 设 T 为 $\mathscr{X}\to\mathscr{T}$ 的变换, A、$B\subset\mathscr{X}$, 则有(i) $A\subset B\Rightarrow T(A)\subset T(B)$; (ii) $T(A\cup B)=T(A)\cup T(B)$; (iii) $T(A\cap B)\subset T(A)\cap T(B)$, 但 $T(A\cap B)\supset T(A)\cap T(B)$ 未必成立. 又若 C、$D\subset\mathscr{T}$, 则有 (iv) $T^{-1}(C\cup D)=T^{-1}(C)\cup T^{-1}(D)$; (v) $T^{-1}(C\cap D)=T^{-1}(C)\cap T^{-1}(D)$.

18. 设 $(\mathscr{X}, \mathscr{B}_x)$ 为一可测空间, T 为 $\mathscr{X}\to\mathscr{T}$ 的变换. 试证明

(i) 若 $\mathscr{B}_T$ 为 $\mathscr{T}$ 上的一 σ-域，则 $T^{-1}(\mathscr{B}_T)$ 亦为 σ-域;

(ii) 若 $\mathscr{A}$ 为 $\mathscr{T}$ 的某些子集构成的类，则有 $T^{-1}(\sigma(\mathscr{A}))=\sigma(T^{-1}(\mathscr{A}))$.

其中记号 $\sigma(\mathscr{A})$ 表示由 $\mathscr{A}$ 产生的最小 σ-域，等等.

(iii) $\{A: A\subset\mathscr{T},\ T^{-1}(A)\in\mathscr{B}_x\}$ 为 σ-域.

19．设样本空间为 $(R^n,\ \mathscr{B}^n)$，统计量的值域空间为 $(R^m,\ \mathscr{B}^m)$．记 $\mathscr{B}_0^n$ 为 R^n 中一切可列集及其余集的全体构成的 σ-域．试证明不存在统计量 T: $(R^n,\ \mathscr{B}^n)\to(R^m,\ \mathscr{B}^m)$，使 $\mathscr{B}_0^n=T^{-1}(\mathscr{B}^m)$.

20．设 T、S 为定义于样本空间 $(\mathscr{X},\ \mathscr{B}_x)$ 上的两个统计量，值域空间分别为 $(\mathscr{T},\ \mathscr{B}_T)$、$(\mathscr{S},\ \mathscr{B}_s)$，且有 $S^{-1}(\mathscr{B}_s)\subset T^{-1}(\mathscr{B}_T)$，还存在变换 φ: $\mathscr{T}\to\mathscr{S}$，使对一切 $x\in\mathscr{X}$，有 $S=\mathscr{S}[T(x)]$．试举例说明 $\mathscr{S}(t)$ 未必是 $(\mathscr{T},\ \mathscr{B}_T)\longrightarrow(\mathscr{S},\ \mathscr{B}_s)$ 可测变换.

21．设 $T: (\mathscr{X},\ \mathscr{B}_x,\ P)\to(\mathscr{T},\ \mathscr{B}_T,\ P^T)$ 是一统计量．$f_1(X)$、$f_2(X)$ 是两个互相独立的它们本身及其积均值存在有限的随机变量．试举例说明：未必有

$$E[f_1(X)f_2(X)\,|\,t]=E[f_1(X)\,|\,t]E[f_2(X)\,|\,t],\ \ a.s.P^T.$$

22．试证下面等式的一边有意义，则另一边也有意义，而且 $a.s.P^T$ 相等：

$$E[f_1(X)E(f_2(X)\,|\,t)]=E[f_2(X)E(f_1(X)\,|\,t)].$$

23．设 $f(X)$ 为均值有限且与 $T(X)$ 独立的随机变量，则有

$$E[f(X)\,|\,t]=E(f(X)),\ \ a.s.P^T.$$

24．设 $T:(\mathscr{X},\ \mathscr{B}_x,\ P)\to(\mathscr{T},\ \mathscr{B}_T,\ P^T)$ 是一统计量，作一新的统计量 $T^*(x)=(x,\ T(x))$，其值域空间为 $(\mathscr{X}\times\mathscr{T},\ \mathscr{B}_x\times\mathscr{B}_T)$．又设 $f(X)$ 为均值存在有限的随机变量．试证明 $E[f(X)\,|\,t^*]=f(X)$，$a.s.P$.

25．设 T_1、T_2 是两个统计量，$f(X)$ 均值有限．试举例说明下式未必成立：

$$E[E(f(X)\,|\,T_1)\,|\,t_2]=E[f(X)\,|\,t_1,\ t_2],\ \ a.s.P^{T_2}.$$

26．设 $T(X)$ 与 $f(X)$ 独立，$f(X)$ 均值有限，则有

$$E[f(X)+T(X)\,|\,t]=t+E[f(X)],\ \ a.s.P^T.$$

27．设 $X_1,\ \cdots,\ X_n$ 为 iid．随机变量．X_1 具有密度 $f(x_1)dx_1$．令 $T=\min\{X_1,\ \cdots,\ X_n\}$，试证明

$$E[X_1\,|\,t]=\frac{t}{n}+\frac{n-1}{n}\int_t^\infty xf(x)dx\Big/\int_t^\infty f(x)dx,\ \ a.s.P^T.$$

28．若 T、S 为定义于 $(\mathscr{X},\ \mathscr{B}_x,\ P_\theta,\ \theta\in\Theta)$ 上的两个统计量，其值域空间 $(\mathscr{T},\ \mathscr{B}_T)$、$(\mathscr{S},\ \mathscr{B}_s)$ 均是欧氏的，则若 $\mathscr{T}$ 充分，且 $T^{-1}(\mathscr{B}_T)\subset T^{-1}(\mathscr{B}_s)$，那么 S 也充分.

29．设 $\{P_\theta,\ \theta\in\Theta\}$ 为样本空间 $(\mathscr{X},\ \mathscr{B}_x)$ 上的一个分布族，设 $Y:(\mathscr{X},\ \mathscr{B}_x)\to(\mathscr{Y},\ \mathscr{B}_y)$ 为一统计量，则 $T(x)=(x,\ y(x))$ 为一充分统计量（设其值域空间为 $(\mathscr{X}\times\mathscr{Y},\ \mathscr{B}_x\times\mathscr{B}_y)$）.

30. 设 $\{P_\theta, \theta\in\Theta\}$ 为 $(\mathscr{X}, \mathscr{B}_x)$ 上的分布族，μ 为其上一σ-有限测度．对每一 $\theta\in\Theta$，有 $f_\theta(x)=dP_\theta(x)/d\mu$．对每一固定的 $x\in\mathscr{X}$，把 $f_\theta(x)$ 看成是 Θ 上的函数，记其为 $t=T(x)$．把这样的元素全体记为 $\mathscr{T}$．于是 $T:\mathscr{X}\to\mathscr{T}$ 是一变换．令 $\mathscr{B}_T=\{A:A\subset\mathscr{T}, T^{-1}(A)\in\mathscr{B}_x\}$．试证明：$T$ 为值域是$(\mathscr{T}, \mathscr{B}_T)$的充分统计量．

31. 设 $X_1, \cdots, X_n$ 为 iid. 随机变量．$X_1\sim N(0, \theta), \theta>0$．直接从定义证明：$T(x)=\sum\limits_1^n x_i^2$ 为充分统计量．

32. 设 $X_1,\cdots,X_n$ 为 iid. 随机变量，$X_1\sim N(a,\sigma^2)$，$-\infty<a<\infty$，$\sigma\geqslant0$．试证明 $(\overline{X}, S^2)$ 仍为充分统计量，这里 $\overline{X}=\frac{1}{n}\sum\limits_{i=1}^n X_i$，$S^2=\sum\limits_1^n(X_i-\overline{X})$ （注意，此时可能 $\sigma=0$，故因子判别定理不能直接应用）．

参 考 文 献

[1] M. Loeve: Probability Theory, Van Nostrand, 1960.

[2] E. L. Lehmann: Testing Statistical Hypothesis, John Wiley, 1959.

[3] H. Cramer: Mathematical Methods of Statistics, 1946.

第二章 无偏估计

设随机变量 X 的样本空间为$(\mathscr{X}, \mathscr{B}_x)$. X 的分布属于 $\mathscr{B}_x$ 上的某个分布族$(P_\theta, \theta\in\Theta)$,即存在 $\theta_0\in\Theta$,使 X 的分布为 P_{θ_0}.参数估计的课题是要通过 X 的观察值即样本值 x, 去估计 θ_0 或某个函数 $g(\theta)$在 θ_0 点的值 $g(\theta_0)$. 通常称 $g(\theta)$为待估函数. 这里的 θ_0 称为"参数真值". 一般为了方便计,常用 θ 代替 θ_0. 因此,θ 这个符号既用来泛指任一可能的参数值(即取值于 Θ 中的流动元),又用来指在一个特定问题中参数的实在值. θ 的这种双重身份,在具体的场合根据上下文的意义一般不会引起混淆,不过,随时注意这一点, 将有助于我们更确切地理解参数估计问题的意义.

按照 A. Wald 的统计判决理论,参数估计问题只不过是具有某些特点的统计判决问题. 对这些特点,难以给出精确的描述,大致上可以归纳为以下几点:

1. 参数空间 Θ 或 $g(\theta)$的值域空间是欧氏空间中的一个区间,或至少是形状比较"规则"的集合.

2. 行动空间 A 与 $g(\theta)$的值域重合,或包含这个值域.

3. 损失函数 $L(\theta, a)$与 $g(\theta)$同 a 的"距离"有密切关系:距离愈大(小),损失愈大(小). 但一般说来,$L(\theta, a)$不一定完全取决于某种距离. 另外,$L(\theta, a)$往往要有一定良好的分析性质,如连续性、可微性、凸性等.

参数的点估计理论是受统计判决思想影响最大的一个统计分支. 但是从历史上看:本章所要讨论的主题——无偏估计的概念,曾经在参数估计的发展上起过重要的作用. 这恐怕主要是由于以下两个原因:一是这个概念有相当的直观背景;一是它在数学上比较好讨论. 但不能不承认,由于判决理论的影响,使人们对无偏性

概念的估价有所改变．这主要是无偏性对于缩小风险未必总是有益的．本章的出发点是把无偏估计作为一个数学对象来讨论．它在某一特定问题中是否合用，应当根据问题的具体性质来分析．

本章在第一节介绍无偏估计的基本概念；第二节引进一个重要概念——分布族与统计量的完全性．并以此为工具，导出本章的主要结果——广义的 Blackwell-Rao-Lehmann-Scheffe 定理，它是寻找一致最优无偏估计的主要方法．以下三节是关于这个基本定理在一些具体类型的分布族中的应用．最后一节是关于线性模型参数的最优线性无偏估计，在那里我们介绍重要的 Gauss-Markov 理论．

§1 基本概念

一、无偏估计的定义

设 X 的样本空间为 $(\mathscr{X}, \mathscr{B}_x)$，分布族为 $(P_\theta, \theta\in\Theta)$，行动空间为 $(A, \mathscr{B}_A)$．$g(\theta)$ 定义于 Θ 上，取值于 R^m．要由 X 的观察值 x 去估计 $g(\theta)$．在参数估计问题中，随机化的判决函数称为随机化估计，非随机化判决函数称为非随机化估计．

我们曾在第一章 1.(三).5 中提到过无偏估计的概念．现将无偏估计的定义正式表述如下：

定义 1.1 设 $\hat{g}(x)$ 为 $g(\theta)$ 的一个估计量．若对任何 $\theta\in\Theta$，$\hat{g}(X)$ 的均值总等于被估计的 $g(\theta)$，则称 $\hat{g}$ 为 $g(\theta)$ 的一个无偏估计．用式子表示，即：

$$E_\theta[\hat{g}(X)]=\int_{\mathscr{X}}\hat{g}(x)dP_\theta(x)=g(\theta),\ \ \theta\in\Theta. \tag{1.1}$$

这是对非随机化估计而言的．对于随机化估计 $\hat{g}(\cdot|x)$，其无偏性类似地定义如下：

$$\int_{\mathscr{X}}\left[\int_A a\hat{g}(da|x)\right]dP_\theta(x)=g(\theta),\ \ \theta\in\Theta.$$

由无偏估计的定义可知：若一估计有无偏性，则将这估计量

多次使用时，各次所得的估计值的平均与被估计的 $g(\theta)$ 很接近。更确切地说：若有 n 个人对 X 进行独立的观察(也可以设想成同一个人在 n 天内每天对 X 独立观察一次，参数真值始终保持为 θ)，第 i 个人得样本 X_i，他可用 $\hat{g}(X_i)$ 去估计 $g(\theta)$。因此，一共得到 $g(\theta)$ 的 n 个估计值 $g(X_i)$ $(i=1, \cdots, n)$。根据强大数定律，当 $n\to\infty$ 时，这 n 个估计值的平均 $\frac{1}{n}\sum_1^n \hat{g}(X_i)$ 以概率为 1 地收敛于 $g(\theta)$。

因此，无偏性的作用在于可以把重复估计中的各次误差通过平均来消除。这当然不意味着在该估计量一次使用时必能获得良好的结果。因此，在具体问题中无偏性是否合理，应当结合具体情况来考虑。在有些问题中，无偏性的要求会导出很不合理的结果来。

例 1.1 设变量 X 服从 Poisson 分布 $\mathscr{P}(\theta)$，即

$$P_\theta(X=k)=\frac{\theta^k}{k!}e^{-\theta} \quad (k=0, 1, 2, \cdots; 0<\theta<\infty).$$

要估计 $g(\theta)=e^{-2\theta}$。若 $\hat{g}(x)$ 为 $g(\theta)$ 的无偏估计，则由定义应有

$$\sum_{k=0}^{\infty}\hat{g}(k)\frac{e^{-\theta}\theta^k}{k!}=e^{-2\theta}, \quad 0<\theta<\infty.$$

即
$$\sum_{k=0}^{\infty}\hat{g}(k)\frac{\theta^k}{k!}=\sum_{k=0}^{\infty}(-1)^k\frac{\theta^k}{k!}, \quad 0<\theta<\infty.$$

由幂级数相等的条件，即得

$$\hat{g}(k)=(-1)^k \quad (k=0, 1, 2, \cdots).$$

这就是说，$g(\theta)$ 的无偏估计只有一个：当观察值为偶数时，利用 1 估计 $e^{-2\theta}$；当观察值为奇数时，利用 -1 来估计 $e^{-2\theta}$。这从哪个角度看都不是一个合理的估计。特别是，$g(\theta)>0$，但估计值有很大可能为负值。

还有这样的情况：无偏估计根本不存在。

例 1.2 设变量 X 服从二项分布 $B(n, \theta)$ $(0\leqslant\theta\leqslant1)$。要估计 $g(\theta)=\frac{1}{1+\theta^2}$。若 $\hat{g}$ 为无偏估计，则有

$$\sum_{0}^{n}\hat{g}(k)\binom{n}{k}\theta^k(1-\theta)^{n-k}=\frac{1}{1+\theta^2}\quad(0\leqslant\theta\leqslant1).$$

但上式左端为 θ 的多项式，而右边却不然，因此二者在区间[0, 1]处处相等是不可能的．这就是说：$g(\theta)$的无偏估计不存在．

二、一致最优的无偏估计

假若待估函数 $g(\theta)$ 具有无偏估计，我们就称其为**可估的**．常常有这样的情况：$g(\theta)$的无偏估计有不止一个．如 $X_1, \cdots, X_n$ 为取自均值为 θ 的母体的 $iid.$ 样本，那么 $\hat{\theta}_1(x)=x_1$, $\hat{\theta}_2(x)=\frac{1}{2n}\left((n+1)x_n+\sum_1^{n-1}x_i\right)$, $\hat{\theta}_3(x)=\bar{x}$ 都是 θ 的无偏估计，而且从直觉上我们感到第三个估计比前两个要好．为了在无偏估计中比较优劣，需要有个标准．为此需要引进损失函数 $L(\theta, a)$，根据第一章1.(二)2，每个估计量都有一定的风险．如在第一章1.(三).5中所述．我们引进如下定义：

定义 1.2 设 $\mathscr{D}_u$ 是 $g(\theta)$ 的一类无偏估计，而 $\hat{g}^*\in\mathscr{D}_u$．若对任何 $\hat{g}\in\mathscr{D}_u$ 及 $\theta\in\Theta$，有 $R(\theta, \hat{g}^*)\leqslant R(\theta, \hat{g})$，则称 $\hat{g}^*$ 为类 $\mathscr{D}_u$ 中(在损失 $L(\theta, a)$下)的**一致最优无偏估计**．其中 $R(\theta, \hat{g}^*)$、$R(\theta, \hat{g})$分别为 $\hat{g}^*$、$\hat{g}$ 的在损失函数 $L(\theta, a)$下的风险．

最常见的情况是：$\mathscr{D}_u$ 为 $g(\theta)$ 的一切非随机化的无偏估计类 $\mathscr{D}_N$ 或一切随机化的无偏估计类(它包含了一切非随机化的无偏估计)$\mathscr{D}_R$ 或一切线性无偏估计类(见§6)，等等．在这三个场合，上述定义中的 $\hat{g}^*$ 分别称为**一致最优的非随机化无偏估计**、**一致最优的随机化无偏估计**、**一致最优的线性无偏估计**。

我们应注意：无偏估计的定义不要求引进任何损失函数，因此也就很自然，无偏性的要求有时可能与"风险尽可能小"这个要求有冲突．如在第一章例1.6中我们曾看到：那里的无偏估计 $\delta(x)$(下节要证明它是最优无偏估计)其风险处处大于非无偏估计 $\delta_1(x)$．这正是从判决函数理论的观点看，无偏性的要求未必总是合理的道理所在。然而，是否完全应该从判决理论的观点去估

价一个统计方法的好坏，在统计界也存在不同的看法. 一般说来，英国学派对这种观点至少是不热心的.

无论从理论或应用的观点来看，凸损失都占有重要地位. 现在我们给出一个正式的定义：

定义1.3 设行动空间为$(R^m, \mathscr{B}^m)$，若对任何$\theta\in\Theta$，损失函数$L(\theta, a)$作为a的函数是R^m上的(严)凸函数，则称L为(严)凸损失函数.

最重要的严凸损失函数是二次损失函数

$$L(\theta, a)=c(\theta)(g(\theta)-a)'B(g(\theta)-a).$$

这里$g(\theta)$、a都视为m维列向量，B为m阶正定阵，而$c(\theta)>0$.

定理1.1 在凸损失下，$\mathscr{D}_N$构成$\mathscr{D}_R$的本质完全类.

证 设$\hat{g}(\cdot|x)$是$g(\theta)$的任一随机化的无偏估计. 定义

$$\hat{g}(x)=\int_A a\hat{g}(da|x),$$

因为$E_\theta\hat{g}(X)=\int_{\mathscr{X}}\left[\int_A a\hat{g}(da|x)\right]dP_\theta(x)=g(\theta)$，故$\hat{g}$为$g(\theta)$的非随机化的无偏估计. 再由Jenson不等式(第一章中的定理3.7)，对任一$x\in\mathscr{X}$，有

$$L(\theta, \hat{g}(x))\leqslant\int_A L(\theta, a)\hat{g}(da|x).$$

从而

$$\begin{aligned}R(\theta, \hat{g})&=\int_{\mathscr{X}}L(\theta, \hat{g}(x))dP_\theta(x)\\&\leqslant\int_{\mathscr{X}}\left[\int_A L(\theta, a)\hat{g}(da|x)\right]dP_\theta(x)=R(\theta, \hat{g}(\cdot|x)).\end{aligned}$$

定理得证.

由本定理可知：在凸损失下只需考虑非随机化的无偏估计就够了. 今后如无特别声明，在提到无偏估计时，总是指非随机化的无偏估计.

当损失函数为严凸时，我们又有如下事实：

定理1.2 在严凸损失下，风险有限的一致最优无偏估计若存在，则必唯一(即对任一$\theta\in\Theta$，在a. s. P_θ意义下唯一).

证 用反证法：若 $g(\theta)$ 存在两个风险有限的一致最优无偏估计 $\hat{g}_1$、$\hat{g}_2$，使存在 $\theta_0\in\Theta$，

$$P_{\theta_0}\{\hat{g}_1(X)\neq\hat{g}_2(X)\}>0. \tag{1.2}$$

记 $\hat{g}_*(x)=\frac{1}{2}[\hat{g}_1(x)+\hat{g}_2(x)]$，显然 $\hat{g}_*$ 也是 $g(\theta)$ 的无偏估计．由损失函数 L 的严凸性及(1.2)式，有

$$P_{\theta_0}\{L(\theta_0,\hat{g}_*(X))<\frac{1}{2}L(\theta_0,\ \hat{g}_1(X))+L(\theta_0,\ \hat{g}_2(X))\}>0.$$

由上式及 $R(\theta_0,\ \hat{g}_1)=R(\theta_0,\ \hat{g}_2)<\infty$，有

$$\begin{aligned}R(\theta_0,\ \hat{g}_*)&=\int_{\mathscr{X}}L(\theta_0,\ \hat{g}_*(x))dP_{\theta_0}(x)\\&<\frac{1}{2}\left\{\int_{\mathscr{X}}L(\theta_0,\ \hat{g}_1(x))dP_{\theta_0}(x)+\int_{\mathscr{X}}L(\theta_0,\ \hat{g}_2(x))dP_{\theta_0}(x)\right\}\\&=\frac{1}{2}(R(\theta_0,\ \hat{g}_1)+R(\theta_0,\ \hat{g}_2))=R(\theta_0,\ \hat{g}_1),\end{aligned}$$

此与 $\hat{g}_1$ 的最优性相矛盾．定理得证．

当 $m=1$，即待估函数 $g(\theta)$ 为一维实函数时，二次损失变为 $L(\theta,\ a)=c(\theta)(g(\theta)-a)^2$．此时，$g(\theta)$ 的无偏估计 $\hat{g}$ 的风险为

$$R(\theta,\ \hat{g})=c(\theta)E_\theta[\hat{g}(X)-g(\theta)]^2=c(\theta)\mathrm{Var}_\theta\hat{g}(X).$$

显然，因为 $c(\theta)>0$，若 $\hat{g}_1$、$\hat{g}_2$ 均无偏时，

$$R(\theta,\ \hat{g}_1)\leqslant R(\theta,\ \hat{g}_2)\Leftrightarrow\mathrm{Var}_\theta\hat{g}_1(X)\leqslant\mathrm{Var}_\theta\hat{g}_2(X).$$

因此在二次损失下的一致最优无偏估计又称为**方差一致最小的无偏估计**(Uniformly Minimum Variance Unbiased Estimate)，简记为 UMVUE．关于 UMVUE，下面给出一个常用的判别条件：

定理 1.3 $g(\theta)$ 的风险有限的无偏估计 $\hat{g}_*(x)$ 为 UMVUE 的充要条件是：对 0 的任一无偏估计 $\hat{\phi}(X)$，使得对任一 $\theta\in\Theta$，假若 $\mathrm{Var}_\theta\hat{\phi}<\infty$，则有

$$\mathrm{cov}_\theta(\hat{g}_*,\ \hat{\phi})=0. \tag{1.3}$$

证 假定条件(1.3)成立．设 $\hat{g}$ 是 $g(\theta)$ 的任一无偏估计．若 $\mathrm{Var}_\theta\hat{g}=\infty$，则有 $\mathrm{Var}_\theta\hat{g}_*\leqslant\mathrm{Var}_\theta\hat{g}$． 若 $\mathrm{Var}_\theta\hat{g}<\infty$，记 $\hat{\phi}=\hat{g}-\hat{g}_*$，

则 $\hat{\phi}$ 为0的无偏估计，且

$$E_\theta\hat{\phi}^2 \leqslant 2E_\theta(\hat{g}-g(\theta))^2+2E_\theta(g(\theta)-\hat{g}_*)^2<\infty.$$

于是由(1.3)式，即得

$$\begin{aligned}E_\theta(\hat{g}-g(\theta))^2 &= E_\theta(\hat{\phi}+\hat{g}_*-g(\theta))^2\\ &= E_\theta\hat{\phi}^2+2\operatorname{cov}_\theta(\hat{\phi},\ \hat{g}_*)+E_\theta(\hat{g}_*-g(\theta))^2\\ &\geqslant E_\theta(\hat{g}_*-g(\theta))^2.\end{aligned}$$

故 $\hat{g}_*$ 为 $g(\theta)$ 的 UMVUE，从而充分性得证．为证必要性，设 $\hat{g}_*$ 为 $g(\theta)$ 的 UMVUE，风险有限．又设 $\hat{\phi}$ 为任一0的无偏估计．对任一 $\theta\in\Theta$，假若 $\mathrm{Var}_\theta\hat{\phi}<\infty$，记之为 a．若 $a=0$，则 $\hat{\phi}(X)=0$，a. s. P_θ. (1.3)式自然成立．故不妨设 $0<a<\infty$．再记 $b=\operatorname{cov}_\theta(\hat{g}_*,\ \hat{\phi})$．假定(1.3)不成立，即 $b\neq 0$．当 $b>0$时，取 $\lambda\in\left(-\frac{2b}{a},0\right)$，当 $b<0$ 时，取 $\lambda\in\left(0,-\frac{2b}{a}\right)$．并记 $\hat{g}_\lambda=\hat{g}_*+\lambda\hat{\phi}$．显然，$\hat{g}_\lambda$ 为 $g(\theta)$ 的无偏估计．但是

$$\mathrm{Var}_\theta\,\hat{g}_\lambda=\mathrm{Var}_\theta\,\hat{g}_*+2\lambda b+\lambda^2 a<\mathrm{Var}_\theta\,\hat{g}_*.$$

这与 $\hat{g}_*$ 的最优性又相矛盾．定理得证．

由此定理又可得到下面简单而有用的事实：

系 1.1 若 $\hat{g}_i$ 为 $g_i(\theta)$ 的风险有限的 UMVUE，c_i、d 为常数 $(i=1,\ \cdots,\ m)$．则 $\sum_1^m c_i\hat{g}_i+d$ 为 $\sum_1^m c_i g_i(\theta)+d$ 的 UMVUE.

证 显然，$\sum_1^m c_i\hat{g}_i+d$ 为 $\sum_1^m c_i g_i(\theta)+d$ 的方差有限的无偏估计．设 $\hat{\phi}$ 是0的任一无偏估计，对任一 $\theta\in\Theta$，若 $\mathrm{Var}_\theta\hat{\phi}<\infty$，由定理1.3的必要性，有

$$\operatorname{cov}_\theta\left(\sum_1^m c_i\hat{g}_i+d,\ \hat{\phi}\right)=\sum_1^m c_i\operatorname{cov}_\theta(\hat{g}_i,\ \hat{\phi})+0=0.$$

再由定理1.3的充分性，此系便得证．

三、无偏估计的协方差阵

设 $g(\theta)=(g_1(\theta),\ \cdots,\ g_m(\theta))'$ 为可估的 m 维向量函数；$\hat{g}(x)=(\hat{g}_1(x),\ \cdots,\ \hat{g}_m(x))'$ 为 $g(\theta)$ 的无偏估计．若对任一 θ 有

$\mathrm{Var}_\theta \hat{g}_i < \infty$ ($i=1, \cdots, m$). 则 $\hat{g}$ 存在协方差阵

$$\mathrm{VAR}_\theta \hat{g} = (\mathrm{cov}_\theta(\hat{g}_i, \hat{g}_j))_{1\leqslant i,j\leqslant m}, \quad \theta \in \Theta,$$

且为非负定. 对任意两个 m 阶非负定阵 A 与 B, 若 $A-B$ 为非负定,则记为 $A \geqslant B$; 若 $A-B$ 为正定,则记为 $A > B$.

假定 $g(\theta)$具有使协方差阵存在的无偏估计. 我们有如下结果:

定理1.4 设 $\hat{g}^* = (\hat{g}_1^*, \cdots, \hat{g}_m^*)'$为 $g(\theta) = (g_1, \cdots, g_m)'$ 的无偏估计. 在二次损失 $L(\theta, a) = (g(\theta) - a)'(g(\theta) - a)$下, $\hat{g}^*$ 为 $g(\theta)$的一致最优无偏估计的充要条件是: $\hat{g}^*$ 的协方差阵存在, 且对 $g(\theta)$的任意协方差阵存在的无偏估计 $\hat{g} = (\hat{g}_1, \cdots, \hat{g}_m)'$ 有

$$\mathrm{VAR}_\theta \hat{g}^* \leqslant \mathrm{VAR}_\theta \hat{g} \quad (\text{对任意 } \theta \in \Theta). \tag{1.4}$$

证 因为 $E_\theta(g(\theta) - \hat{g}^*)'(g(\theta) - \hat{g}^*) = \mathrm{tr}(\mathrm{cov}_\theta \hat{g}^*)$; $E_\theta(g(\theta) - \hat{g})'(g(\theta) - \hat{g}) = \mathrm{tr}(\mathrm{cov}_\theta \hat{g})$. 此处 $\mathrm{tr}(A)$表示方阵 A 的迹, 又因为 $A \geqslant B \Rightarrow \mathrm{tr}A \geqslant \mathrm{tr}B$, 故由(1.4)式可推出 $\hat{g}^*$ 为 $g(\theta)$的一致最优无偏估计.

反之,若在二次损失下, $\hat{g}^*$ 为 $g(\theta)$的一致最优无偏估计; $\hat{g}$ 为 $g(\theta)$ 的任一协方差阵存在的无偏估计. 则因 $\sum\limits_{j=1}^m \mathrm{Var}_\theta \hat{g}_j^* \leqslant \sum\limits_{j=1}^m \mathrm{Var}_\theta \hat{g}_j < \infty$, 故 $\hat{g}^*$ 的协方差阵存在. 又可证明 $\hat{g}_i^*$ 为 $g_i(\theta)$的 UMVUE ($i=1, \cdots, m$). 因为若不然,则存在 i,不妨设 $i=1$,使 $g_1(\theta)$ 有另一无偏估计 $\hat{g}_1^0(x)$, 对某θ, 有 $\mathrm{Var}_\theta \hat{g}_1^0 < \mathrm{Var}_\theta \hat{g}_1^*$. 令 $\hat{g}^0 = (\hat{g}_1^0, \hat{g}_2^*, \cdots, \hat{g}_m^*)'$,则也为 $g(\theta)$的无偏估计,但却有

$$\sum_2^m \mathrm{Var}_\theta(\hat{g}_i^*) + \mathrm{Var}_\theta \hat{g}_1^0 < \sum_1^m \mathrm{Var}_\theta \hat{g}_i^*.$$

此与 $\hat{g}^*$ 的最优性相矛盾. 既然 $\hat{g}_i^*$ 为 $g_i(\theta)$的 UMVUE($i=1, \cdots, m$), 由系 1.1 知: $\sum\limits_1^m c_i \hat{g}_i^*$ 为 $\sum\limits_1^m c_i g_i(\theta)$ 的 UMVUE, 此处 $C = (c_1, \cdots, c_m)'$为任一 m 维常向量. 于是

$$C'(\mathrm{VAR}_\theta(\hat{g}^*))C = \mathrm{Var}_\theta\Big(\sum_1^m c_i \hat{g}_i^*\Big) \leqslant \mathrm{Var}_\theta\Big(\sum_1^m c_i \hat{g}_i\Big) = C'(\mathrm{VAR}_\theta(\hat{g}))C.$$

即有 $\mathrm{VAR}_\theta(\hat{g}^*)\leqslant\mathrm{VAR}_\theta(\hat{g})$. 定理证毕.

以上，我们是在假定(在凸损失或严凸损失下)一致最优无偏估计存在的前提下，讨论了它的若干性质. 下面几节我们要讨论一致最优无偏估计的存在性及其求解方法.

§2 完全统计量

一、基本概念

为研究一致最优无偏估计的问题，我们引入分布族与统计量完全性的概念. 设 $\mathscr{P}=(P_\theta,\ \theta\in\Theta)$为样本空间$(\mathscr{X},\ \mathscr{B}_x)$上的分布族. $g(\theta)$是可估函数，$\hat{g}_0(x)$为其无偏估计，$\mathbf{0}$ 是维数与 $g(\theta)$ 相同的零向量. 以 $\mathscr{D}_0$ 表示 $\mathbf{0}$ 的无偏估计的全体. 于是 $g(\theta)$的无偏估计全体可表示为

$$\mathscr{D}=\{\hat{g}_0+\hat{\phi}:\hat{\phi}\in\mathscr{D}_0\}.$$

因此，对 $g(\theta)$的无偏估计的研究可转变为对 $\mathbf{0}$ 的无偏估计的研究. 特别是，当 $\mathbf{0}$ 的无偏估计唯一时，$\hat{g}_0$ 是 $g(\theta)$的唯一的无偏估计. 此时一致最优无偏估计的求解问题就变得很简单了. 这样，我们给出分布族和统计量的完全性的概念:

定义 2.1 (i) 若对任一 $\mathscr{P}$ 可积函数 $\phi(x)$，由"对任一 $\theta\in\Theta$ 有 $\int_{\mathscr{X}}\phi(x)dP_\theta(x)=0$"可推出"对任一 $\theta\in\Theta$ 有 $\phi(X)=0$, a.s. P_θ". 则称 $\mathscr{P}$ 是完全分布族.

(ii) 设 T: $(\mathscr{X},\ \mathscr{B}_x,\ P_\theta,\ \theta\in\Theta)\to(\mathscr{T},\ \mathscr{B}_T,\ P_\theta^T,\ \theta\in\Theta)$ 是一统计量，若分布族 $\mathscr{P}^T=\{P_\theta^T,\ \theta\in\Theta\}$ 完全，则称 T 为完全统计量.

从定义即知：若 T 完全，则对任一 $\theta\in\Theta$ 依赖于 T 的 $\mathbf{0}$ 的无偏估计 a.s. P_θ^T 为 0. 从而可估函数 $g(\theta)$的基于 T 的无偏估计必唯一. 由此可见：完全统计量特别是完全充分统计量在无偏估计理论中占有特殊重要的地位.

例 2.1 二项分布族 $B(n,\ \theta)$ $(0<\theta<1)$是完全的.

证 对任一实函数 $\phi(x)$，若对所有的 $\theta\in(0,\ 1)$有

$$\sum_{x=0}^{n}\phi(x)\binom{n}{x}\theta^x(1-\theta)^{n-x}=0.$$

令 $t=\theta/(1-\theta)$，于是对任一 $t\in(0,\ \infty)$ 有

$$\sum_{x=0}^{n}\left[\phi(x)\binom{n}{x}\right]t^x=0.$$

故对 $x=0,\ \cdots,\ n$，有 $\phi(x)\binom{n}{x}=0$，从而 $\phi(x)=0$. 由定义知：二项分布族是完全的.

例 2.2 正态分布族 $N(a,\ \sigma^2)$ $(-\infty<a<\infty;\ 0<\sigma)$ 是完全的.

证 设 $\phi(x)$ 对所有的 a、σ 有

$$\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}\phi(x)\exp\left\{-\frac{1}{2\sigma^2}(x-a)^2\right\}dx=0.$$

从而

$$\int_{-\infty}^{\infty}\phi(x)\exp\left\{-\frac{1}{2\sigma^2}x^2+\frac{a}{\sigma^2}x\right\}dx=0.$$

固定 $\sigma=1$，对任意 a，有

$$\int_{-\infty}^{\infty}e^{ax}\left[\phi(x)\exp\left(-\frac{1}{2}x^2\right)\right]dx=0.$$

可见 0 是 $\phi(x)e^{-\frac{1}{2}x^2}$ 的双边 Laplace 变换，故有

$$\phi(x)e^{-\frac{1}{2}x^2}=0,\ \text{a. e. L}.$$

即对所有的 a、σ 有 $\phi(X)=0$, a. s. $N(a,\ \sigma^2)$. 从而正态分布族是完全的.

利用类似的方法，不难证明常见的分布族，如 Poisson 分布族 $\mathscr{P}(\lambda)$ $(0<\lambda<\infty)$；均匀分布族 $R(0,\ \theta)$ $(0<\theta<\infty)$；Γ 分布族 $G(\alpha,\ p)$ $(0<\alpha<\infty;\ 0<p<\infty)$；$\beta$ 分布族 $B(p,\ q)$ $(0<p<\infty,\ 0<q<\infty)$；负指数分布族 $dp(x;\ \alpha,\ \beta)=\frac{1}{\beta}e^{-\frac{1}{\beta}(x-\alpha)}dx$ $(x\geqslant\alpha,\ \beta>0,\ -\infty<\alpha<\infty)$，等等都是完全的. 但不完全的分布族也是常见的.

例 2.3 设 $X=(X_1,\ \cdots,\ X_n)'$ 的分布为 $N(a\mathbb{1}_n,\ \sigma^2I_n)$

$(-\infty<a<\infty,\ 0<\sigma)$. 这个分布族是不完全的. 事实上, 记 $\phi(x)=x_1-x_2$. 易验证: 对任意的 a、σ, 有 $E_{(a,\sigma)}\phi(X)=0$. 但 "$\phi(X)=0$, a. s. $N(a\mathbb{1}_n,\ \sigma^2 I_n)$" 并不成立. 此处及以下, $\mathbb{1}_n$ 表示用 n 个 1 组成的列向量.

不过, 在此例中, 若令 $T(x)=\bar{x}$, 因 $\bar{X}=\frac{1}{n}\sum_1^n X_i$ 的分布为 $N\left(a,\frac{1}{n}\sigma^2\right)$ $(-\infty<a<\infty,\ \sigma>0)$. 由例 2.2 知: $\bar{X}$ 是完全统计量.

例 2.4 设 $X_1,\ \cdots,\ X_n$ 为 iid. 随机变量, $X_1\sim R(0,\ \theta)$ $(0<\theta<\infty)$. 那么 $T(x)=\max(x_1,\ \cdots,\ x_n)$ 是完全充分统计量. 在第一章例 6.5 中已证明了 T 的充分性. 为证其完全性, 注意 $T(X)$ 的分布为

$$dP_\theta^T(t)=\frac{n}{\theta^n}t^{n-1}dt\quad(0<t<\theta).$$

若 $\phi(t)$ 对任意的 $\theta>0$ 有 $\int_0^\theta\phi(t)\left(\frac{n}{\theta^n}\right)t^{n-1}dt=0$, 即对任意的 $\theta>0$ 有

$$\int_0^\theta\phi(t)t^{n-1}dt=0.$$

于是根据实变函数论的有关论述, 得

$$\phi(t)t^{n-1}=0\quad \text{a. e. L 于 }(0,\ \infty).$$

因此对任意 $\theta\in(0,\ \infty)$ 有 $\phi(t)=0$, a. s. P_θ^T. 这就证明了 T 的完全性.

关于分布族和统计量的完全性有下面简单的事实:

定理 2.1 (i)设 $\mathscr{P}'$ 是分布族 $\mathscr{P}$ 的子族, 若 $\mathscr{P}'$ 完全, 且 $\mathscr{P}'\sim\mathscr{P}$. 则 $\mathscr{P}$ 也完全. (ii)设 T: $(\mathscr{X},\ \mathscr{B}_x,\ P_\theta,\ \theta\in\Theta)\to(\mathscr{T},\ \mathscr{B}_T,\ P_\theta^T,\ \theta\in\Theta)$ 是一统计量, 若 $(P_\theta,\ \theta\in\Theta)$ 完全, 则 T 也完全.

证 (i)若对任一 $P\in\mathscr{P}$ 有 $\int_{\mathscr{X}}\phi(x)dP(x)=0$, 因为 $\mathscr{P}'\subset\mathscr{P}$, 所以对任一 $P\in\mathscr{P}'$ 有 $\int_{\mathscr{X}}\phi(x)dP=0$. 由 $\mathscr{P}'$ 的完全性, 对任一 $P\in\mathscr{P}'$ 有 $P\{\phi(X)\neq0\}=0$. 又因 $\mathscr{P}'$ 与 $\mathscr{P}$ 等价. 于是对任一 $P\in\mathscr{P}$ 也有 $P\{\phi(X)\neq0\}=0$. 于是(i)得证.

(ii) 若$\int_{\mathscr{T}}\phi(t)dP_\theta^T(t)=0$对每一$\theta\in\Theta$成立，从而对所有的$\theta\in\Theta$，有$\int_{\mathscr{X}}\phi(T(x))dP_\theta(x)=0$. 因为$\{P_\theta,\ \theta\in\Theta\}$完全，则对任意$\theta\in\Theta$，有$0=P_\theta\{\phi(T(x))\neq0\}=P_\theta^T\{\phi(t)\neq0\}$. 于是(ii)得证.

由例2.3知：此定理的(ii)的逆不成立.

由定理的(i)立即得出如下结果：若$\mathscr{P}'$、$\mathscr{P}$为同一样本空间上的两个等价分布族，且$\mathscr{P}'\subset\mathscr{P}$. 若统计量$T$在分布族$\mathscr{P}'$下是完全的，那么它在分布族$\mathscr{P}$下也是完全的. 由定理的(ii)，又可推出下面的结果：若T、S为在同一样本空间上定义的两个统计量，其值域空间$(\mathscr{T},\ \mathscr{B}_T)$、$(S,\ \mathscr{B}_S)$都是欧氏的；且若$T^{-1}(\mathscr{B}_T)\subset T^{-1}(\mathscr{B}_S)$，且$S$是完全的，则$T$也是完全的. 这由第一章定理4.1及本定理立即可得. 可见，完全统计量经过进一步"加工"简化后，仍然是完全的. 完全性的这一特性与充分性正好相反(参看第一章习题28).

下面我们引入所谓"严格完全性"的概念，利用这一概念，可以得出一个关于乘积分布族完全性的判别定理.

定义2.2 设$\mathscr{P}=(P_\theta,\ \theta\in\Theta)$为$(\mathscr{X},\ \mathscr{B}_x)$上的一个分布族，$\mu$为参数空间$(\Theta,\ \mathscr{B}_\Theta)$上的一个$\sigma$-有限测度. 对每一$B\in\mathscr{B}_x$，$P_\theta(B)$关于$\mathscr{B}_\Theta$可测. 假若对任一$\phi(x)$，由条件"对任一给定的$\mu$零集$N$，对每一$\theta\overline{\in}N$，有$\int\phi(x)dP_\theta=0$"可推出"对所有的$\theta\in\Theta$，有$\phi(X)=0$, a. s. P_θ."则称$\mathscr{P}$为严格完全的.

不言而喻，严格完全是一种更强的完全性. 容易验证，例2.1、2.2及其后面所列出的诸分布族都是严格完全的(取L测度作为定义2.2中的μ).

定理2.2 设$\mathscr{P}=(P_\theta,\ \theta\in\Theta)$是$(\mathscr{X},\ \mathscr{B}_x)$上的完全分布族，$\widetilde{\mathscr{P}}=(\widetilde{P}_{\tilde\theta},\ \tilde\theta\in\widetilde\Theta)$为$(\widetilde{\mathscr{X}},\ \mathscr{B}_{\tilde x})$上关于$(\widetilde\Theta,\ \mathscr{B}_{\tilde\Theta},\ \mu)$是严格完全分布族. 则乘积分布族

$$\mathscr{P}\times\widetilde{\mathscr{P}}=(P_\theta\times\widetilde{P}_{\tilde\theta},\ \theta\in\Theta,\ \tilde\theta\in\widetilde\Theta)$$

为完全的.

证 设函数 $\phi(x, \tilde{x})$ 使对每一 $(\theta, \tilde{\theta})\in\Theta\times\tilde{\Theta}$ 有

$$\int_{\mathscr{X}\times\tilde{\mathscr{X}}}\phi(x, \tilde{x})dP_\theta(x)d\tilde{P}_{\tilde{\theta}}(\tilde{x})=0. \tag{2.1}$$

记 $C=\{(x, \tilde{x}): \phi(x, \tilde{x})\neq 0\}$，以下要证：对每一 $(\theta, \tilde{\theta})\in\Theta\times\tilde{\Theta}$，有

$$P_\theta\times\tilde{P}_{\tilde{\theta}}(C)=\int_{\mathscr{X}}\tilde{P}_{\tilde{\theta}}(C_x)dP_\theta(x)=0, \tag{2.2}$$

其中 C_x 表示集 C 在 x 处的截口.

由条件(2.1)及 Fubini 定理，对每一 $\tilde{\theta}\in\tilde{\Theta}$ 有

$$\int_{\mathscr{X}}\left[\int_{\tilde{\mathscr{X}}}\phi(x, \tilde{x})d\tilde{P}_{\tilde{\theta}}(\tilde{x})\right]dP_\theta(x)=0 \quad (\text{对所有的 } \theta\in\Theta).$$

由 $\mathscr{P}$ 的完全性，对每一 $\theta\in\Theta$，有

$$\int_{\tilde{\mathscr{X}}}\phi(x, \tilde{x})d\tilde{P}_{\tilde{\theta}}(\tilde{x})=0, \text{ a. s. } P_\theta. \tag{2.3}$$

记 $B=\{(x, \tilde{\theta}): \int_{\tilde{\mathscr{X}}}\phi(x, \tilde{x})d\tilde{P}_{\tilde{\theta}}(\tilde{x})\neq 0\}$. 由 $\tilde{P}_{\tilde{\theta}}$ 关于 $\mathscr{B}_{\tilde{\Theta}}$ 的可测性知 B 关于 $\mathscr{B}_x\times\mathscr{B}_{\tilde{\Theta}}$ 可测. 因 B 在 $\tilde{\theta}$ 处截口为

$$B_{\tilde{\theta}}=\{x: \int_{\tilde{\mathscr{X}}}\phi(x, \tilde{x})d\tilde{P}_{\tilde{\theta}}(\tilde{x})\neq 0\},$$

由(2.3)知：对每一 $\theta\in\Theta$，有 $P_\theta(B_{\tilde{\theta}})=0$. 那么，在乘积空间 $(\mathscr{X}\times\tilde{\Theta}, \mathscr{B}_x\times\mathscr{B}_{\tilde{\Theta}}, P_\theta\times\mu)$ 上，有

$$(P_\theta\times\mu)(B)=\int_{\tilde{\Theta}}P_\theta(B_{\tilde{\theta}})d\mu(\tilde{\theta})=0.$$

再由 Fubini 定理知：对每一 $\theta\in\Theta$，有

$$\int_{\mathscr{X}}\mu(B^x)dP_\theta(x)=0.$$

此处 B^x 表示集 B 在 x 处的截口. 因为 $\mu(B^x)\geqslant 0$，故对每一 $\theta\in\Theta$，有 $\mu(B^x)=0$, a. s. P_θ. 于是存在 P_θ 零集 N_θ，对每一 $\theta\in\Theta$，使当 $x\bar{\in}N_\theta$ 时，$\mu(B^x)=0$. 由 B 的定义，当 $\tilde{\theta}\bar{\in}B^x$ 时，有

$$\int_{\tilde{\mathscr{X}}}\phi(x, \tilde{x})d\tilde{P}_{\tilde{\theta}}(\tilde{x})=0,$$

从而由 $\tilde{\mathscr{P}}$ 的严格完全性，对每一 $\tilde{\theta}\in\tilde{\Theta}$，有

$$\phi(x, \tilde{x})=0, \text{ a. s. } \tilde{P}_{\tilde{\theta}}.$$

这就是说：对每一 $x\bar{\in}N_\theta$，$\tilde{\theta}\in\tilde{\Theta}$，存在 $\tilde{P}_{\tilde{\theta}}$ 零集 $N_{(\tilde{\theta}, x)}$，使当

$\tilde{x}\overline{\in} N_{(\theta,x)}$ 时，$\phi(x,\tilde{x})=0$. 因而，
当 $x\overline{\in} N_\theta$ 时，有 $C_x\subset N_{(\tilde{\theta},x)}$. 最后有

$$0\leqslant(P_\theta\times\widetilde{P}_{\tilde{\theta}})(C)=\int_{\mathscr{X}-N_\theta}P_{\tilde{\theta}}(C_x)dP_\theta(x)$$
$$\leqslant\int_{\mathscr{X}-N_\theta}P_{\tilde{\theta}}(N_{(\tilde{\theta},x)})dP_\theta(x)=0.$$

定理证毕.

二*、次序统计量的完全性

本小节证明在广泛的分布族中，次序统计量为完全充分统计量.

设 $\mathscr{P}=\{F\}$ 是任一给定的一维分布族，$X_1,\cdots,X_n$ 为从分布 $F\in\mathscr{P}$ 中抽得的简单样本，而 $(X_{(1)},\cdots,X_{(n)})$ 为次序统计量. $\varphi(x_1,\cdots,x_n)$ 为任一 $F\times\cdots\times F(F\in\mathscr{P})$ 可积函数，那么由第一章例 5.1 知:

$$E_F(\varphi(X_1,\cdots,X_n)\,|\,x_{(1)},\cdots,x_{(n)})$$
$$=\frac{1}{n!}\sum\nolimits_{(i_1,\cdots,i_n)}\varphi(x_{i_1},\cdots,x_{i_n})\triangleq U,\tag{2.4}$$

其中 $(i_1,\cdots,i_n)$ 为 $(1,\cdots,n)$ 的任一排列. 可见次序统计量为分布族 $(F\times\cdots\times F:F\in\mathscr{P})$ 的充分统计量.

Frasor 于 1954 年证明了：在相当广泛的一类分布族中，次序统计量也是完全的. 下面我们采用陈希孺的证法，这样可以得到更一般性的结果.

定理 2.3 设 $\mathscr{P}$ 为一个一维分布族，满足条件：(i) $\mathscr{P}$ 为凸集，即若 F、$G\in\mathscr{P}$，则 $pG+qF\in\mathscr{P}$，此处 $0\leqslant p\leqslant1$，$q=1-p$; (ii) 若 $F\in\mathscr{P}$，$-\infty<a<b<\infty$，且 $F\{[a,b)\}>0$，则 $F_{[a,b)}\in\mathscr{P}$，其中 $F_{[a,b)}$ 定义为

$$F_{[a,b)}(A)=F\{A\cap[a,b)\}/F\{[a,b)\}.$$

则次序统计量 $T(X)=(X_{(1)},\cdots,X_{(n)})$ 关于分布族 $\{F\times\cdots\times F:F\in\mathscr{P}\}$ 为完全的.

证 首先指出，由 (i) 知，若 $F_i\in\mathscr{P}$，$p_i\geqslant0$ $(i=1,\cdots,k)$ $\sum_1^k p_i=1$，则有 $\sum_1^k p_iF_i\in\mathscr{P}$. 这由

$$p_1F_1+p_2F_2+p_3F_3=(p_1+p_2)\left[\frac{p_1}{p_1+p_2}F_1+\frac{p_2}{p_1+p_2}F_2\right]+p_3F_3\in\mathscr{P}$$

可以看出.

设任一 $F\in\mathscr{P}$，$E_F(g(T(X_1,\cdots,X_n)))=0$，即有

$$\int\cdots\int_{R_n}g(T(x_1,\cdots,x_n))dF(x_1)\cdots dF(x_n)=0,\tag{2.5}$$

在下面证明中，随时注意有$g(T(x_1, \cdots, x_n))=g(T(x_{i_1}, \cdots, x_{i_n}))$，此处$(i_1, \cdots, i_n)$为$(1, \cdots, n)$的任一排列.

设$\{i_1, \cdots, i_k\}$为$\{1, \cdots, n\}$的任一非空子集，定义

$I(i_1, \cdots, i_k)=\{(j_1, \cdots, j_n): j_1, \cdots, j_n \in \{i_1, \cdots, i_k\}$，且$(i_1, \cdots, i_k)$中每一元在$(j_1, \cdots, j_n)$中至少出现一次$\}$. 以下证明

$$C_k(i_1, \cdots, i_k) = \sum_{(j_1, \cdots, j_n)\in I(i_1, \cdots, i_k)} \int \cdots \int_{R^n} g(T(x_1, \cdots, x_n)) dF_{j_1}(x_1) \cdots dF_{j_n}(x_n) = 0, \tag{2.6}$$

这里$F_1, \cdots, F_n \in \mathscr{P}$.

用数学归纳法：当$k=1$时，不失一般性，可令$i_1=1$，(2.6)式成为

$$\int \cdots \int_{R^n} g(T(x_1, \cdots, x_n)) dF_1(x_1) \cdots dF_1(x_n)=0,\ F_1 \in \mathscr{P}.$$

这就是(2.5)式.

设$k \leqslant m-1$时，(2.6)式成立，记$\bar{F}=\frac{1}{m}\sum_1^m F_i \in \mathscr{P}$. 由(2.5)式知

$$\int \cdots \int_{R^n} g(T(x_1, \cdots, x_n)) d\bar{F}(x_1) \cdots d\bar{F}(x_n)=0.$$

把上式左边展开，得

$$\frac{1}{m^n}\Big\{\sum_{1\leqslant j_1\leqslant m} C_1(j_1)+\sum_{1\leqslant j_1<j_2\leqslant m} C_2(j_1, j_2)+\cdots$$
$$+\sum_{1\leqslant j_1<\cdots<j_{m-1}\leqslant m} C_{m-1}(j_1, \cdots, j_{m-1})+C_m(1, \cdots, m)\Big\}=0.$$

由归纳假定，便推得$C_m(1, \cdots, m)=0$. 同样对任意的$1\leqslant i_1<i_2<\cdots<i_m\leqslant n$，也有$C_m(i_1, \cdots, i_m)=0$. 从而(2.6)式对$k=m$也成立.

特别取$n=m$时，对任意的$F_1, \cdots, F_n \in \mathscr{P}$有

$$\int \cdots \int_{R^n} g(T(x_1, \cdots, x_n)) dF_1(x_1) \cdots dF_n(x_n)=0. \tag{2.7}$$

现取n个区间$[a_1, b_1), \cdots, [a_n, b_n)$，对任一分布$F \in \mathscr{P}$，可证

$$\int \cdots\cdots \int_{[a_1, b_1)\times\cdots\times[a_n, b_n)} g(T(x_1, \cdots, x_n)) dF(x_1) \cdots dF(x_n)=0. \tag{2.8}$$

事实上，若有某个i，使$F\{[a_i, b_i)\}=0$，(2.8)式显然成立. 若对每一个i，$F\{[a_i, b_i)\}>0$ $(i=1, \cdots, n)$. 令$F_i=F_{[a_i, b_i)}$，由条件(ii)知：$F_i \in \mathscr{P}$ $(i=1, \cdots, n)$. 把诸F_i代入(2.7)式，便推出(2.8)式.

令$\mu(A)=\int \cdots \int_A g(T(x)) dF(x_1) \cdots dF(x_n)$，$A \in \mathscr{B}^n$，则$\mu$为$\mathscr{B}^n$上的有号

测度，但在一切形如 $A=[a_1, b_1)\times\cdots\times[a_n, b_n)$ 的集合上，$\mu(A)=0$，故由测度扩张定理，在 $\mathscr{B}^n$ 上 $\mu\equiv 0$，即对一切 $A\in\mathscr{B}^n$，有

$$\int\cdots\int_A g(T(x))dF(x_1)\cdots dF(x_n)=0.$$

从而，对一切 $F\in\mathscr{P}$，推得

$$F\times\cdots\times F(x: g(T(x))\neq 0)=0.$$

这就证明了 T 的完全性. 定理证毕.

此定理可应用于下面几种情况：$\mathscr{P}$ 为 1°. 一切一维分布族；2°. 一切关于 L 测度绝对连续的一维分布族；3°. 一切离散型的一维分布族；4°. 一切 $r(>0)$ 阶矩阵存在有限的一维分布族. 在这些分布族中，次序统计量都是完全充分的.

§3 一致最优无偏估计的存在性

一、广义 Blackwell-Rao-Lehmann-Scheffe 定理

由充分统计量的充分性原则(第一章定理 6.3)知：若 T 是充分统计量，那么在凸损失下的一致最优无偏估计可以在依赖于 T 的无偏估计类中寻找. 如若 T 还是完全的，则由上节所述，基于 T 的无偏估计是唯一的. 从而我们可以看出这样一个重要事实：依赖于一个完全充分统计量的无偏估计必为一致最优无偏估计. 这一事实是我们综合了 Blackwell 和 Rao 以及 Lehmann 和 Scheffé 的工作而自然得到的结果，简称为广义 B-R-L-S 定理.

定理 3.1 设 $g(\theta)$ 为一可估函数，则有：(i) 在凸损失下，依赖于充分统计量 T 的无偏估计的全体构成本质完全类；(ii) 在凸损失下，若 T 是完全充分统计量，则一致最优无偏估计必存在；在严凸损失下，若 $g(\theta)$ 至少存在一个有限风险的无偏估计，则依赖于完全充分统计量 T 的无偏估计，必是唯一的一致最优无偏估计.

证 (i) 由第一章引理 6.1：存在与 θ 无关的条件概率分布 $P(B|t)$, $B\in\mathscr{B}_x$, $t\in\mathscr{T}$. 若 $\hat{g}(x)$ 为 $g(\theta)$ 的任一无偏估计，令

$$\hat{g}_*(t)=E[\hat{g}(X)|t]=\int_{\mathscr{X}}\hat{g}(x)P(dx|t), \tag{3.1}$$

因为

$$E_\theta \hat{g}_*(T(X))=\int_{\mathscr{X}}\hat{g}_*(T(x))dP_\theta(x)=\int_{\mathscr{T}}\hat{g}_*(t)dP_\theta^T(t)$$

$$=\int_{\mathscr{T}}\left[\int_{\mathscr{X}}\hat{g}(x)P(dx|t)\right]dP_\theta^T(t)=\int_{\mathscr{X}}\hat{g}(x)dP_\theta(x)=g(\theta),$$

所以 $\hat{g}_*(T(x))$ 为 $g(\theta)$ 的无偏估计. 再由 Jensen 不等式, 得

$$R(\theta,\ \hat{g}_*)=\int_{\mathscr{X}}L(\theta,\ \hat{g}_*(T(x)))dP_\theta(x)$$

$$=\int_{\mathscr{T}}L(\theta,\ \hat{g}_*(t))dP_\theta^T(t)$$

$$\leqslant\int_{\mathscr{T}}\left[\int_{\mathscr{X}}L(\theta,\ \hat{g}(x))P(dx|t)dP_\theta^T(t)\right.$$

$$=\int_{\mathscr{X}}L(\theta,\ \hat{g}(x))dP_\theta(x)=R(\theta,\ \hat{g}).$$

于是(i)得证.

(ii)由 T 的充分性, 利用(3.1)式可求得 $g(\theta)$ 的依赖于 T 的无偏估计 $\hat{g}_*(T(x))$. 又由 T 的完全性知, $\hat{g}_*(T(x))$ 必为 $g(\theta)$ 的唯一的依赖于 T 的无偏估计. 但由于(i), 单元素集 $\{\hat{g}_*(T)\}$ 构成本质完全类, 故 $\hat{g}_*$ 为一致最优.

若再假定损失是严凸的, 且 $g(\theta)$ 存在有限风险的无偏估计, 由 $\hat{g}_*$ 的最优性知, 其风险有限, 再根据定理 1.2, $\hat{g}_*$ 必为 $g(\theta)$ 的唯一的一致最优无偏估计. 定理证毕.▌

广义 B-R-L-S 定理是无偏估计理论中最重要的基本结果. 它的意义在于为我们提供了一个求解一致最优无偏估计的有效方法: 如果损失是严凸的, 首先找一个完全充分统计量, 然后设法求解依赖于 T 的无偏估计, 从而得到要求的解. 而求解依赖于 T 的无偏估计的常用的方法是: 先找任一无偏估计 $\hat{g}(x)$, 再计算条件期望 $E[\hat{g}|t]$ 即可.

另一方面, 广义 B-R-L-S 定理还告诉我们这样一个事实: 当完全充分统计量存在时, 在各种不同的严凸损失下, 一致最优无偏估计的解是相同的. 因此, 我们只需考虑二次损失就可以了. 特别当 $g(\theta)$ 是实函数时, 我们将着重求解其 UMVUE.

二、若干例子

下面举几个例子，以说明广义 B-R-L-S 定理的初步应用.

例 3.1 设 $X_1, \cdots, X_n$ 为 $iid.$ 随机变量，$X_1 \sim \mathscr{P}(\lambda)$, $0<\lambda<\infty$. 现在来估计 $g(\lambda)=P_\lambda[X_1=k]=e^{-\lambda}\lambda^k/k!$.

因为 $(X_1, \cdots, X_n)$ 的联合分布为

$$P_\lambda[X_1=x_1, \cdots, X_n=x_n]=e^{-n\lambda}\lambda^{\sum_1^n x_i}\Big/\prod_1^n x_i!$$

$$(x_i=0, 1, \cdots; i=1, \cdots, n).$$

由因子判别定理(第一章中的定理 6.1)，$T=\sum_1^n x_i$ 为充分统计量. 又因 $T\sim\mathscr{P}(n\lambda)$, $0<\lambda<\infty$. 故 T 是完全的. 当 $x_1=k$ 或其它时，令 $\hat{g}(x)=1$、0. 它是 $g(\lambda)$ 的一个无偏估计. 当 $T=t$ 时，X 的条件分布为

$$\begin{aligned}
&P_\lambda\left[X_1=i\,\Big|\,\sum_1^n X_j=t\right]\\
&\quad=P_\lambda[X_1=i]P_\lambda\left[\sum_1^n X_j=t\,\Big|\,X_1=i\right]\Big/P_\lambda\left[\sum_1^n X_j=t\right]\\
&\quad=\binom{t}{i}\left(\frac{1}{n}\right)^i\left(1-\frac{1}{n}\right)^{t-i}\ (\text{当 } i\leqslant t).
\end{aligned}$$

故 $\quad E(\hat{g}\,|\,t)=P[X_1=k\,|\,t]=\binom{t}{k}\left(\frac{1}{n}\right)^k\left(1-\frac{1}{n}\right)^{t-k}\quad(k\leqslant t).$

根据广义 B-R-L-S 定理，要求的 UMVUE 为

$$\hat{g}_*(t)=\begin{cases}\binom{t}{k}\left(\frac{1}{n}\right)^k\left(1-\frac{1}{n}\right)^{t-k}, & \text{当 } k\leqslant t;\\ 0, & \text{当其它}.\end{cases}$$

再者，因 $E_\lambda(\overline{X})=\frac{1}{n}\sum_{i=1}^n E(X_i)=\lambda$, 故 $\overline{X}$ 为 λ 的 UMVUE.

例 3.2 设 $X_1, \cdots, X_n$ 为 $iid.$ 随机变量，$X_1\sim G(x_1; 1/\theta, p)$ $(0<\theta<\infty)$, 自由度 p 已知. 估计可靠性函数 $R_\tau(\theta)=P_\theta[X_1\geqslant\tau]$, τ 是已知常数.

(i) 先证 $T=\sum_1^n x_i$ 是完全充分统计量．因为 $(X_1, \cdots, X_n)$ 的联合密度为

$$\prod_{i=1}^n G(x_i; 1/\theta, p)=\frac{1}{\theta^{np}[\Gamma(p)]^n}\left(\prod_{i=1}^n x_i\right)^{p-1} e^{-\frac{1}{\theta}\sum_1^n x_i}$$

$$(0<x_i; i=1, \cdots, n).$$

故由因子判别定理知，T 是充分的．又因 $T=\sum_1^n X_i \sim G(t; 1/\theta, np)$，$0<\theta<\infty$，故 T 又是完全的．

(ii) $I_{[X_1\geqslant\tau]}(X)$ 是 $R_\tau(\theta)$ 的无偏估计．现在求 $P[X_1\geqslant\tau|t]$．

注意：X_1 与 $Y_{n-1}=\sum_2^n X_i$ 是独立的，而 $Y_{n-1}\sim G(y; 1/\theta, (n-1)p)$，则 (X_1, Y_{n-1}) 的联合密度为

$$f_{X_1, Y_{n-1}}(x, y; \theta)=G(x; 1/\theta, p)G(y; 1/\theta, (n-1)p)$$

$$=\frac{1}{\theta^p\Gamma(p)}x^{p-1}e^{-\frac{x}{\theta}}\frac{1}{\theta^{(n-1)p}\Gamma((n-1)p)}y^{(n-1)p-1}e^{-y/\theta}$$

$$(x\geqslant 0, y\geqslant 0).$$

作变换 $T=X_1+Y_{n-1}$，$X_1=X_1$，得 (X_1, T) 的联合密度为

$$f_{X_1, T}(x, t; \theta)=[\theta^{np}\Gamma(p)\Gamma((n-1)p)]^{-1}x^{p-1}(t-x)^{(n-1)p-1}e^{-t/\theta}$$

$$(0\leqslant x\leqslant t<\infty),$$

故在 $T=t$ 条件下，X_1 的条件密度为

$$f_{X_1|T}(x, t; \theta)=\frac{\Gamma(np)}{\Gamma(p)\Gamma((n-1)p)}\left(\frac{1}{t}\right)\left(\frac{x}{t}\right)^{p-1}\left(1-\frac{x}{t}\right)^{(n-1)p-1}$$

$$(0\leqslant x\leqslant t<\infty),$$

从而得到 $R_\tau(\theta)$ 的 UMVUE 为

$$P[X_1\geqslant\tau|t]=\begin{cases}0, & t<\tau\\ \displaystyle\int_\tau^t\frac{\Gamma(np)}{\Gamma(p)\Gamma((n-1)p)} \\ \qquad\displaystyle\times\left(\frac{1}{t}\right)\left(\frac{x}{t}\right)^{p-1}\left(1-\frac{x}{t}\right)^{(n-1)p-1}dx, & t\geqslant\tau\end{cases}$$

$$=\begin{cases}0, & t<\tau;\\ \displaystyle\int_{\tau/t}^1[B(p, (n-1)p)]^{-1}z^{p-1}(1-z)^{(n-1)p-1}dz, & t\geqslant\tau.\end{cases}$$

特别是当 $p=1$ 时，UMVUE 为 $\hat{R}_\tau=0$; $(1-t/\tau)^{n-1}$，当 $t<\tau$ 或 $t\geqslant\tau$ 时.

例 3.3 设 $X_1, \cdots, X_n$ 为 iid. 随机变量，X 具有负指数分布

$$f(x_1;\alpha,\beta)=\begin{cases}\dfrac{1}{\beta}\exp\left[-\dfrac{x_1-\alpha}{\beta}\right], & x_1\geqslant\alpha\\ 0, & x_1<\alpha\end{cases}$$

$$(-\infty<\alpha<\infty,\ \beta>0),$$

$(X_{(1)}, \cdots, X_{(n)})$为次序统计量. 现利用前 r 个观察值$(x_{(1)}, \cdots, x_{(n)})$来估计可靠性函数

$$R_\tau(\alpha,\beta)=P_{\alpha,\beta}[X_1\geqslant\tau]$$

$$=e^{-\frac{1}{\beta}(\tau-\alpha)}\quad 或\quad 1\ (当\ \tau\geqslant\alpha\quad 或\quad \tau<\alpha).$$

由第一章例 6.6 知$(X_{(1)}, T)$是充分统计量，其中 $T=\sum\limits_{2}^{r}X_{(i)}+(n-r)X_{(r)}-(n-1)X_{(1)}$. 又因为$(X_{(1)}, \cdots, X_{(r)})$的分布密度为

$$f(x_{(1)},\cdots,x_{(r)};\alpha,\beta)$$

$$=\begin{cases}\dfrac{n!}{(n-r)!}\beta^{-r}\exp\left\{-\dfrac{\sum\limits_{1}^{r}x_{(i)}+(n-r)x_{(r)}}{\beta}+\dfrac{n\alpha}{\beta}\right\}, & \alpha\leqslant x_{(1)};\\ 0, & \alpha>x_{(1)}.\end{cases}$$

作线性变换 $X_{(1)}=X_{(1)}$, $T=\sum\limits_{2}^{r}X_{(i)}+(n-r)X_{(r)}-(n-1)X_{(1)}$, $Y_i=(n-i+1)(X_{(i)}-X_{(i-1)})\ (i=3, \cdots, r)$，可推得$(X_{(1)}, T, Y_3, \cdots, Y_r)$的密度为

$$g(x_{(1)}, t, y_3, \cdots, y_r;\alpha,\beta)$$

$$=\begin{cases}n\beta^{-r}\exp\left\{-\dfrac{t}{\beta}-\dfrac{n}{\beta}(x_{(1)}-\alpha)\right\},\\ \qquad 当\ \alpha\leqslant x_{(1)},\ 0\leqslant t,\ y_i\geqslant 0\ (i=3,\cdots,r),\ y_3+\cdots+y_r\leqslant t;\\ 0,\quad 对其它情况.\end{cases}$$

于是$(X_{(1)}, T)$的分布密度为

$$g(x_{(1)},t;\alpha,\beta)=\int_{y_i>0\ (i=3,\cdots,r),\ \sum\limits_{3}^{r} y_i<t} g(x_{(1)},t,y_3,\cdots,y_r;\alpha,\beta)dy_3\cdots dy_r$$

$$=\frac{n}{\beta}e^{-\frac{n}{\beta}(x_{(1)}-\alpha)}\frac{1}{\Gamma(r-1)\beta^{r-1}}t^{r-2}e^{-\frac{t}{\beta}},$$

$$t>0,\ x_{(1)}\geqslant\alpha,\begin{pmatrix}-\infty<\alpha<\infty;\\0<\beta<\infty\end{pmatrix}. \tag{3.2}$$

利用定理 2.2 的证明方法,不难推证(3.2)族是完全的. 从而证明了$(X_{(1)}, T)$为完全充分统计量.

下面来验证 $R_\tau(\alpha, \beta)$的 UMVUE 为

$$\hat{R}_\tau(X_{(1)}, T)=\begin{cases}\dfrac{n-1}{n}\left[\max\left(0,\left(1-\dfrac{\tau-X_{(1)}}{T}\right)\right)\right]^{r-2}, \\ \qquad\qquad\qquad 当\ \tau\geqslant X_{(1)}; \\ 1, \qquad\qquad\quad 当\ \tau<X_{(1)}.\end{cases} \tag{3.3}$$

由广义 B-R-L-S 定理, 仅需验证 $E_{(\alpha,\beta)}\hat{R}_\tau(X_{(1)}, T)=R_\tau(\alpha, \beta)$. 根据(3.2)与(3.3)式, 当 $\tau<\alpha$ 时, 有

$$E_{(\alpha,\beta)}\hat{R}_\tau=\int_\alpha^\infty dx\int_0^\infty g(x,\ t;\alpha,\ \beta)dt=1. \tag{3.4}$$

而当 $\tau\geqslant\alpha$ 时, 有

$$E_{(\alpha,\beta)}\hat{R}_\tau=\int_\alpha^\tau dx\int_0^\infty \hat{R}_\tau(x,\ t)g(x,\ t;\alpha,\ \beta)dt$$

$$+\int_0^\infty dx\int_0^\infty \hat{R}_\tau(x,\ t)g(x,\ t;\alpha,\ \beta)dt\triangleq I_1+I_2,$$

其中

$$I_2=\int_t^\infty \frac{n}{\beta}e^{-\frac{n}{\beta}(x-\alpha)}dx\int_0^\infty\frac{1}{\beta^{r-1}}\cdot\frac{1}{\Gamma(r-1)}t^{r-2}e^{-\frac{t}{\beta}}dt$$

$$=e^{-\frac{n(\tau-\alpha)}{\beta}};$$

$$I_1=\int_\alpha^\tau dx\int_0^\infty\left(\frac{n-1}{n}\right)\left[\max\left(0,\left(1-\frac{\tau-x}{t}\right)\right)\right]^{r-2}$$

$$\times\frac{n}{\beta^r}\cdot\frac{1}{\Gamma(r-1)}t^{r-2}e^{-\frac{t}{\beta}-\frac{n(x-\alpha)}{\beta}}dt$$

$$=\int_\alpha^\tau\frac{n-1}{\beta}e^{-\frac{n(x-\alpha)}{\beta}}dx\int_{\tau-x}^\infty\frac{1}{\beta^{r-1}\Gamma(r-1)}$$

$$\times\left(1-\frac{\tau-x}{t}\right)^{r-2}t^{r-2}e^{-\frac{t}{\beta}}dt.$$

令 $z=t-\tau+x$，得

$$I_1=e^{\frac{n\alpha-\tau}{\beta}}\int_\alpha^\tau \frac{n-1}{\beta}e^{-\frac{n-1}{\beta}x}dx\int_0^\infty \frac{1}{\beta^{r-1}}\cdot\frac{1}{\Gamma(r-1)}z^{r-2}e^{-\frac{z}{\beta}}dz$$

$$=e^{-\frac{1}{\beta}(\tau-\alpha)}-e^{-\frac{n(\tau-\alpha)}{\beta}}.$$

从而有

$$E_{(\alpha,\beta)}\hat{R}_\tau(X_{(1)},\ T)=I_1+I_2=e^{-\frac{1}{\beta}(\tau-\alpha)}. \tag{3.5}$$

合(3.4)与(3.5)，便证得 $E_{(\alpha,\beta)}\hat{R}_\tau(X_{(1)},\ T)=R_\tau(\alpha,\ \beta)$.

由这几个例子可见：利用 B-R-L-S 定理，困难往往在于求找完全充分统计量和计算条件期望，而且随着分布族和待估函数的不同，困难程度也各异．关于完全充分统计量和 UMVUE 的结果，可查看[12]中 p.153～p.174.

三、一点附注

在本小节我们要指出：广义 B-R-L-S 定理给出的关于 UMVUE 存在性的条件并不是必要条件.

例 3.4 设 X 的分布为

$$P_\theta[X=k]=\begin{cases}\theta, & \text{当 } k=-1\\ (1-\theta)^2\theta^k, & \text{当 } k=0,\ 1,\ 2,\ \cdots\end{cases}\quad(0\leqslant\theta\leqslant1). \tag{3.6}$$

记 $g_1(\theta)=\theta$, $g_2(\theta)=(1-\theta)^2$. 以下证明 $g_1(\theta)$ 不存在 UMVUE，而 $g_2(\theta)$ 却存在.

设 $\phi(x)$ 是 0 的任一无偏估计．因为对任一 $\theta\in[0,\ 1]$，有

$$0=E_\theta\phi(X)=\phi(-1)\theta+\sum_{k=0}^\infty\phi(k)(1-\theta)^2\theta^k$$

$$=\phi(0)+\sum_{k=1}^\infty(\phi(k-2)-2\phi(k-1)+\phi(k))\theta^k,$$

故有 $\phi(k)=k\phi(1)$ $(k=-1,\ 0,\ 1,\ \cdots)$．若记 $\phi(1)=c$，则 $\phi(x)=cx$. 于是便得到 0 的无偏估计的全体为

$$\mathscr{D}_0=\{cX:\ c\ \text{是实数}\}.$$

当 $X=-1$ 或 $\neq-1$ 时，令 $T_0(X)=1$ 或 0. 容易验证 $T_0(X)$ 是，

$g_1(\theta)$的一个无偏估计. 则 $g_1(\theta)=\theta$ 的无偏估计的全体为

$$\mathscr{D}=\{T_0+cX: c \text{ 是实数}\}.$$

对任一 c, 因为 X 是 0 的无偏估计, 而

$$\operatorname{cov}_\theta(T_0+cX,\ X)=E_\theta[T_0(X)X+cE_\theta X^2-\theta E_\theta X]$$

$$=-\theta+c(\theta+(1-\theta)^2\sum_{k=1}^{\infty}k^2\theta^k)=\theta\left[-1+c\left(1+\frac{1+\theta}{1-\theta}\right)\right].$$

显然, 对固定的 c, 总可以找到 $\theta\in[0,\ 1]$, 使上式 $\neq 0$. 于是由定理 1.3 知, T_0+cX 不为 θ 的 UMVUE. 再由 c 的任意性, 这就证明了 $g_1(\theta)$的 UMVUE 不存在. 又: 当 $x=0$ 时, 令 $T_1(x)=1$; 当 $x=-1,\ 1,\ 2,\ \cdots$ 时, 令 $T_1(x)=0$. 容易验证 $T_1(X)$为 $g_2(\theta)$ 的无偏估计. 又因为对任一实数 c, 有

$$\operatorname{cov}_\theta(T_1,\ cX)=cE_\theta(T_1(X)X)=0,$$

再根据定理 1.3, 即知 $T_1(X)$为 $g_2(\theta)$的 UMVUE.

因为可估函数 $g_1(\theta)$的 UMVUE 不存在, 那么由广义 B-R-L-S 定理可知, 分布族(3.6)不存在完全充分统计量. 但另一方面, $g_2(\theta)$的 UMVUE 却存在, 这说明了完全充分统计量的存在并不是 UMVUE 存在的必要条件. 这就是说, 当完全充分统计量不存在时, 一个可估函数的 UMVUE 可能存在也可能不存在. 在这种情形下, 要判断 UMVUE 存在与否, 乃至当已知其存在时来解它, 就不能简单地套用广义 B-R-L-S 定理了. 此时, 只能根据面临的分布族及待估函数的具体特点, 采用各种不同的特殊方法和技巧来研究解决. 关于这方面的工作可参看 [11]、[15] 等.

但是, 在常见的分布族中, 大部分具有完全充分统计量. 因此, 广义 B-R-L-S 定理得到广泛的应用. 以下几节, 我们将介绍这个定理在几种重要的分布族中的应用.

§4 指数型分布族中的参数无偏估计

本节将讨论广义 B-R-L-S 定理在指数型分布族中的参数无偏中的应用.

一、指数族的完全充分统计量

为了利用 B-R-L-S 定理求解指数族中的 UMVUE，我们先介绍一个重要的完全充分统计量.

定理 4.1 设 X 服从指数型分布族

$$dP_\theta(x)=c(\theta)\exp\left\{\sum_{j=1}^{k}\theta_jT_j(x)\right\}d\mu(x)\quad(\theta\in\Theta).\tag{4.1}$$

假若自然参数空间 Θ 含有内点，则统计量 $T(x)=(T_1(x),\cdots,T_k(x))$ 是完全充分的.

证 由(4.1)式及因子判别定理即知 T 是充分的. 以下证明其完全性：记 $d\mu^T(t)$ 为 $d\mu(x)$ 通过 T 在 $(\mathscr{T},\mathscr{B}_T)$ 上的导出测度，那么概率分布族通过 T 在 $(\mathscr{T},\mathscr{B}_T)$ 上的导出分布族为

$$dP_\theta^T(t)=c(\theta)\exp\left\{\sum_{j=1}^{k}\theta_jt_j\right\}d\mu^T(t)\quad(\theta\in\Theta).\tag{4.2}$$

不失一般性，可设原点为 Θ 的内点. 因若不然，作变换 $\tilde{\theta}=\theta-\theta_0$，其中 θ_0 为 Θ 一内点. 于是新参数空间 $\tilde{\Theta}=\{\tilde{\theta}:\tilde{\theta}=\theta-\theta_0,\ \theta\in\Theta\}$ 以原点为内点. 我们可对新的参数 $\tilde{\theta}$ 来讨论. 从而，存在 $\delta>0$，使 Θ 包含矩形 $I=\{(\theta_1,\cdots,\theta_k):|\theta_j|<\delta,\ j=1,\cdots,k\}$.

设 $f(t)$ 是任一 $\{P_\theta^T,\ \theta\in\Theta\}$ 可积函数，使对任一 $\theta\in\Theta$，有

$$c(\theta)\int_{\mathscr{T}}f(t)e^{\sum_{j=1}^{k}\theta_jt_j}d\mu^T(t)=0.\tag{4.3}$$

因为 $f(t)=f^+(t)-f^-(t)$，f^+、f^- 分别为 f 的正、负部. 于是对每一 $\theta\in\Theta$ 有

$$\int_{\mathscr{T}}e^{\sum_{j=1}^{k}\theta_jt_j}f^+(t)d\mu^T(t)=\int_{\mathscr{T}}e^{\sum_{j=1}^{k}\theta_jt_j}f^-(t)d\mu^T(t).\tag{4.4}$$

根据指数族性质(第一章中的定理 2.2)：(3.4)两边在区域

$\{(\xi_1+i\eta_1,\cdots,\xi_k+i\eta_k):(\xi_1,\cdots,\xi_k)\in I,\ \eta_1,\cdots,\eta_k$ 为实数$\}$

对每一分量 θ_j 是解析函数. 因此，根据复变函数解析延拓定理，有

$$\int_{\mathscr{T}}e^{i\sum_{1}^{k}\eta_jt_j}f^+(t)d\mu^T(t)=\int_{\mathscr{T}}e^{i\sum_{1}^{k}\eta_jt_j}f^-(t)d\mu^-(t).\tag{4.5}$$

在(4.4)式中取 $\theta=0$, 得

$$a\triangleq\int_{\mathscr{T}}f^{+}(t)d\mu^{T}(t)=\int_{\mathscr{T}}f^{-}(t)d\mu^{T}(t)<\infty.$$

若 $a=0$, 则 $f(t)=0$, a. e. μ^{T}, 从而对每一 $\theta\in\Theta$, 有 $f(t)=0$, a. s. P_{θ}^{T}. 若 $a>0$, 记

$$dP^{+}(t)=\frac{1}{a}f^{+}(t)d\mu^{T}(t),dP^{-}(t)=\frac{1}{a}f^{-}(t)d\mu^{T}(t),$$

它们是$(\mathscr{T},\mathscr{B}_{T})$上的两个概率分布. 由(4.5)式知, 这两个概率分布有相同的特征函数, 从而有 $dP^{+}(t)=dP^{-}(t)$, 即有

$f(t)=f^{+}(t)-f^{-}(t)=0$, a. e. μ^{T}, 从而 a. s. P_{θ}^{T} $(\theta\in\Theta)$.

这证明了分布族(3.2)的完全性即 T 的完全性. 定理得证.

由此定理给出的完全充分统计量, 利用 B-R-L-S 定理, 可直接求解指数族的一些较简单的 UMVUE.

例4.1 设 $X_1,\cdots,X_n$ 为 $iid.$ 随机变量, $P_{\theta}[X_1=1]=\theta$, $P_{\theta}[X_1=0]=1-\theta$ $(\theta\in(0,1))$. $P_{\theta}\{X_1=x_1,\cdots,X_n=x_n\}=\theta^{\sum_1^n x_i}(1-\theta)^{n-\sum_1^n x_i}=c(\theta)e^{\varphi T(x)}$, 其中 $c(\theta)=(1-\theta)^n$, $\varphi=\log\left(\frac{\theta}{1-\theta}\right)$ $(-\infty<\varphi<\infty)$, $T(x)=\sum_1^n x_i$. 这是以 φ 为自然参数的指数族. 因为自然参数空间为实数轴, 故由定理4.1知, T 为完全充分统计量. 现要估计$\theta=\frac{e^{\varphi}}{1+e^{\varphi}}$, 记为 $g(\varphi)$. 令 $\hat{g}(T(x))=\frac{1}{n}T(x)=\bar{x}$, 因 $E_{\theta}\left[\frac{1}{n}\sum_1^n X_i\right]=\theta$, 故由B-R-L-S定理知, $\bar{X}$ 为 θ 的 UMVUE.

例4.2 设 $X_1,\cdots,X_n$ 为 $iid.$ 随机变量, $X_1\sim N(a,\sigma^2)$ $(-\infty<a<\infty;\ 0<\sigma<\infty)$. 要估计 $g_1=a$, $g_2=\sigma^2$, $g_3=\frac{a}{\sigma}$, $g_4=P_{(a,\sigma)}[X_1<u]=\Phi\left(\frac{u-a}{\sigma}\right)$, 其中 u 为已知常数.

因为$(X_1,\cdots,X_n)$的分布密度为

$$\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{1}{2\sigma^2}\sum_1^n(x_i-a)^2}dx_1\cdots dx_n$$

$$=c(\theta_1,\theta_2)e^{\theta_1T_1(x)+\theta_2T_2(x)}d\mu(x), \tag{4.6}$$

其中 $\theta_1=\frac{a}{\sigma^2},\ \theta_2=-\frac{1}{2\sigma^2},\ T_1(x)=\sum_1^n x_i,\ T_2(x)=\sum_1^n x_i^2,$

$$c(\theta_1,\ \theta_2)=\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{na^2}{2\sigma^2}},\ d\mu(x)=dx_1\cdots dx_n.$$

自然参数空间 $\Theta=\{(\theta_1,\ \theta_2):\ -\infty<\theta_1<\infty,\ -\infty<\theta_2<0\}$，因其有内点，根据定理 4.1，$(T_1,\ T_2)$为完全充分统计量，于是与其等价的统计量$(\bar{x},\ s^2)=\left(\frac{1}{n}\sum_1^n x_i,\ \sum_1^n(x_i-\bar{x})^2\right)$也是完全充分的. 因容易验证 $\hat{g}_1=\bar{x},\ \hat{g}_2=\frac{1}{n-1}s^2,\ \hat{g}_3=\bar{x}/c_ns\left(c_n=\Gamma\left(\frac{n-1}{2}\right)\Big/\sqrt{2}\,\Gamma\left(\frac{n}{2}\right)\right)$ 分别为 a、σ^2、a/σ 的无偏估计，从而由 B-R-L-S 定理知，它们分别是被估函数的 UMVUE.

下面来计算 $\Phi\left(\frac{u-a}{\sigma}\right)$的 UMVUE. 因为 $I_{[x_1<u]}(x)$是 $\Phi\left(\frac{u-a}{\sigma}\right)$ 的无偏估计，故由 B-R-L-S 定理，$P[X_1<u|\bar{x},\ s^2]$为 $\Phi\left(\frac{u-a}{\sigma}\right)$ 的 UMVUE. 作正交变换

$$Y_1=\sqrt{n}\,\bar{X},\ Y_2=\sqrt{\frac{n}{n-1}}(X_1-\bar{X}),\ Y_i=\sum_{j=1}^n C_{ij}X_j \quad (i=3,\ \cdots,\ n). \tag{4.7}$$

因为 Y_1 与 Y_2、$S^2=\sum_2^n Y_i^2$ 独立，故有条件密度

$$f(y_2|\bar{x},\ s^2)=\frac{f(y_2,\ \bar{x},\ s^2)}{f(\bar{x},\ s^2)}=\frac{f_1(\bar{x})f_3(y_2)f_4(s^2|y_2)}{f_1(\bar{x})f_2(s^2)} =\frac{f_3(y_2)f_4(s^2|y_2)}{f_2(s^2)}, \tag{4.8}$$

其中 $f_3(y_2)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{y_2^2}{2\sigma^2}}$.

$$f_2(s^2)=\begin{cases}\left[2^{\frac{n+1}{2}}\Gamma\left(\frac{n-1}{2}\right)\sigma^2\right]^{-1}\left(\frac{s^2}{\sigma^2}\right)^{\frac{n-1}{2}-1}e^{-\frac{1}{2}\frac{s^2}{\sigma^2}}, & \text{当 } s^2\geqslant 0;\\ 0, & \text{当 } s^2<0.\end{cases}$$

又因为 $Z\triangleq\frac{S^2-Y_2^2}{\sigma^2}\sim\chi^2(n-2)$，$Z$ 与 Y_2 独立，而 $S^2=Y_2^2+\sigma^2 Z$，

故当 $Y_2=y_2$ 时，S^2 的条件密度为 $\sigma^2 Z+y_2^2$ 的密度，即为

$$f_4(s^2|y_2)$$

$$=\begin{cases}\left[2^{\frac{n-1}{2}}\Gamma\left(\frac{n-2}{2}\right)\sigma^2\right]^{-1}\left(\frac{s^2-y_2^2}{\sigma^2}\right)^{\frac{n-2}{2}-1}\exp\left\{-\frac{s^2-y_2^2}{2\sigma^2}\right\}, & \\ \qquad\qquad\qquad\qquad s^2\geqslant y_2^2; \\ 0, \qquad\qquad\qquad\qquad s^2<y_2^2.\end{cases}$$

把以上各式代入(4.8)式，得

$$f(y_2|\bar{x},\ s^2)=\begin{cases}\dfrac{\Gamma\left(\frac{n-1}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{n-2}{2}\right)\sqrt{s^2}}\left(1-\dfrac{y_2^2}{s^2}\right)^{\frac{n-2}{2}-1}, \\ \qquad\qquad\qquad 当\ s^2\geqslant y_2^2; \\ 0, \qquad\qquad\qquad 当\ s^2<y_2^2.\end{cases}$$

于是

$$P[X_1<u|\bar{x},\ s^2]=\int_{\left\{|y_2|<\sqrt{s^2},\ y_2<\sqrt{\frac{n}{n-1}}(u-\bar{x})\right\}}f(y_2|\bar{x},\ s^2)dy_2,$$

令 $z=\frac{y_2}{\sqrt{s^2}}$，得

$$P[X_1<u|\bar{x},\ s^2]$$

$$=\int_{\left\{|z|<1,\ z<\sqrt{\frac{n}{n-1}}\frac{u-\bar{x}}{\sqrt{s^2}}\right\}}\frac{\Gamma\left(\frac{n-1}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{n-2}{2}\right)}(1-z^2)^{\frac{n-4}{2}}dz.$$

利用勒让德等式 $\Gamma(2p)=\frac{1}{\sqrt{\pi}}2^{2p-1}\Gamma(p)\Gamma\left(p+\frac{1}{2}\right)$，有

$$P(X_1<u|\bar{x},\ s^2)$$

$$=\int_{\left\{|z|<1,\ z<\sqrt{\frac{n}{n-1}}\frac{u-\bar{x}}{\sqrt{s^2}}\right\}}\frac{\Gamma(n-2)}{2^{n-3}\Gamma^2\left(\frac{n-2}{2}\right)}(1-z^2)^{\frac{n-4}{2}}dz,$$

记 $v=\sqrt{\frac{n}{n-1}}\frac{u-\bar{x}}{\sqrt{s^2}}$，并作变换 $t=\frac{1-z}{2}$，得

$$P(X_1<u|\bar{x}, s^2)$$

$$=\int_C \frac{1}{B\left(\frac{n-2}{2}, \frac{n-2}{2}\right)}(1-t)^{\frac{n-2}{2}-1} t^{\frac{n-2}{2}-1} dt,$$

其中

$$C=\{t: |1-2t|\leqslant 1,\ 1-2t\leqslant v\}$$

$$=\begin{cases} \phi, & 当\ v<-1; \\ \left\{t: \frac{1-v}{2}<t\leqslant 1\right\}, & 当\ -1\leqslant v\leqslant 1; \\ \{t: 0\leqslant t\leqslant 1\}, & 当\ v>1. \end{cases}$$

于是求得 $\Phi\left(\frac{u-a}{\sigma}\right)$ 的 UMVUE 为

$$\hat{g}_*(\bar{x}, s^2)=\begin{cases} 0, & 当\ v<-1; \\ \int_{\frac{1-v}{2}}^{1} \beta\left(t; \frac{n-2}{2}, \frac{n-2}{2}\right)dt, & 当\ -1\leqslant v\leqslant 1; \\ 1, & 当\ 1<v. \end{cases} \tag{4.9}$$

它的值可通过不完全 β 积分表查得. 此例可参看[3]、[5].

二、单参数指数族参数的 UMVUE

我们考虑单参数指数族

$$dP_\theta(x)=c(\theta)e^{-\theta T(x)}d\mu(x) \quad (\theta\in\Theta), \tag{4.10}$$

其中 $d\mu(x)=h(x)dx$ 关于 L 测度绝对连续, Θ 为 R^1 中的某一开区间. 由定理 4.1, $T(x)$ 为完全充分统计量. 设 $d\mu(x)$ 通过 T 在 $(\mathscr{T}, \mathscr{B}_T)$ 上的导出测度为 $d\mu^T(t)$, 且关于 L 测度绝对连续, $d\mu^T(t)=\nu(t)dt$, 那么 dP_θ 通过 T 在 $(\mathscr{T}, \mathscr{B}_T)$ 上导出的概率测度为

$$dP_\theta^T(t)=c(\theta)e^{-\theta t}\gamma(t)dt \quad (\theta\in\Theta). \tag{4.11}$$

假定 $\varphi(t)$ 是 $g(\theta)$ 的无偏估计, 于是有

$$g(\theta)=c(\theta)\int e^{-\theta t}\gamma(t)\varphi(t)dt \quad (\theta\in\Theta). \tag{4.12}$$

可见 $g(\theta)/c(\theta)$ 为 $\gamma(t)\varphi(t)$ 的双边拉普拉斯变换．根据双边拉氏变换的有关论述，我们有如下结果：

定理 4.2 在上述记号与假定下，又设

(i) $g(s)$ 在区域 $\{s:s=\theta+iu,\ \theta\in\Theta\}$ 上解析；

(ii) 统计量 T 的特征函数 $\phi_\theta(u)=\int e^{iut}c(\theta)\gamma(t)dt$ 满足条件：

(a) $\lim\limits_{|u|\to\infty}\phi_\theta(-u)g(\theta+iu)=0$，在 Θ 的每一闭子区间上一致成立；

(b) $\int_{-\infty}^{\infty}|\phi_\theta(-u)g(\theta+iu)|du<\infty$．

则下式有意义：

$$W(t)=\lim_{M\to\infty}\frac{1}{2\pi i}\int_{c-iM}^{c+iM}\left(\frac{g(s)}{c(s)}\right)e^{ts}\,ds\quad(c\in\Theta,t\in R^1). \tag{4.13}$$

又若

$$\{t:W(t)\neq0\}\subset\{t:\gamma(t)>0\},\ \text{a. e. L}. \tag{4.14}$$

则 $g(\theta)$ 具有依赖于 t 的无偏估计

$$\varphi(t)=\begin{cases}W(t)/\gamma(t), & \text{当 } \gamma(t)>0,\\ 0, & \text{当 } \gamma(t)=0.\end{cases} \tag{4.15}$$

它就是 $g(\theta)$ 的 UMVUE．

为了给出定理的证明，我们先回顾一下有关双边拉氏变换的知识：设 $\varphi(x)$ 是 R^1 上的 L 可测函数，若复变函数

$$\Phi(s)=\int_{-\infty}^{\infty}\varphi(x)e^{-sx}\,dx \tag{4.16}$$

在某一平面区域内有定义，则称其为 $\varphi(x)$ 的双边拉氏变换，并记为 $B(\varphi;s)$．可以证明，若 $\Phi(s_1)$ 与 $\Phi(s_2)$ 存在，$Re(s_1)<Re(s_2)$，则在区域 $\{s:Re(s_1)\leqslant Re(s)\leqslant Re(s_2)\}$ 内 $\Phi(s)$ 均存在．

双边拉氏变换有如下反演公式：设 $\Phi(s)=B(\varphi;s)$ 在 $\{s:\sigma_1<Re(s)<\sigma_2\}$ 内存在，则对任一 $c\in(\sigma_1,\sigma_2)$，有

$$\varphi(x)=\lim_{T\to\infty}\frac{1}{2\pi i}\int_{c-iT}^{c+iT}\Phi(s)e^{sx}\,ds,\quad \text{a. e. L}. \tag{4.17}$$

双边拉氏变换有如下表示定理;

若 $\Phi(s)$ 在 $\{s:\sigma_1<\mathrm{Re}(s)<\sigma_2\}$ 内解析，并满足条件：

(a) 对任意的 $\delta_1<\delta_2$, $[\delta_1, \delta_2]\subset(\sigma_1, \sigma_2)$，有

$$\lim_{s\to\infty}\Phi(s)=0,\ \text{且在}\{s:\delta_1\leqslant\mathrm{Re}(s)\leqslant\delta_2\}\text{内一致成立;} \tag{4.18}$$

(b) 对任一 $c\in(\sigma_1, \sigma_2)$，有

$$\int_{-\infty}^{\infty}|\Phi(c+iu)|du<\infty. \tag{4.19}$$

则存在 R^1 上的 L 可测函数 $\varphi(x)$，使 $B(\varphi; s)$ 的存在范围包含区域 $\{s:\sigma_1<\mathrm{Re}(s)<\sigma_2\}$，而且在此区域内，$B(\varphi; s)\equiv\Phi(s)$.

定理 4.2 的证明 由指数族性质(第一章中的定理 2.2)及假设条件(i)，$g(s)/c(s)$ 在区域 $\{s:s=\theta+iu,\ \theta\in\Theta\}$ 内解析. 又因为

$$g(\theta+iu)/c(\theta+iu)=g(\theta+iu)\int_{-\infty}^{\infty}e^{-(\theta+iu)t}\gamma(t)dt$$

$$=\frac{g(\theta+iu)}{c(\theta)}\phi_\theta(-u),$$

故由假设条件(a)、(b)，有

$$\lim_{|u|\to\infty}\frac{g(\theta+iu)}{c(\theta+iu)}=0,\ \text{且在}\ \Theta\ \text{的每一闭子区间上一致成立, 且}$$

$$\int_{-\infty}^{\infty}\left|\frac{g(\theta+iu)}{c(\theta+iu)}\right|du<\infty \quad (\theta\in\Theta).$$

从而由表示定理知 $\frac{g(s)}{c(s)}$ 存在双边拉氏逆变换 $W(t)$，如(4.13)所示. 由双边拉氏变换的定义，有

$$\frac{g(s)}{c(s)}=\int_{-\infty}^{\infty}e^{-st}W(t)dt,\ s\in\{s: s=\theta+iu,\ \theta\in\Theta\}.$$

从而有

$$g(\theta)=c(\theta)\int_{-\infty}^{\infty}e^{-\theta t}W(t)dt \quad (\theta\in\Theta). \tag{4.20}$$

又由(4.14)及上式，我们有

$$g(\theta)=c(\theta)\int_{\{W(t)\neq0\}}e^{-\theta t}W(t)dt=c(\theta)\int_{\{\gamma(t)>0\}}e^{-\theta t}W(t)dt$$

$$=c(\theta)\int_{\{\gamma(t)>0\}}e^{-\theta t}\left(\frac{W(t)}{\gamma(t)}\right)\gamma(t)dt=c(\theta)\int_{-\infty}^{\infty}\varphi(t)\gamma(t)dt.$$

再由 B-R-L-S 定理，定理 4.2 即得证.

下面一些例子是定理 4.2 的应用.

例 4.3 在单参数指数族(4.11)中，估计 $g(\theta)=\theta^k$，k 为自然数. 假定 $m\leqslant k$，$r^{(m)}(t)$，a. e. L 存在，且 $e^{-\theta t}\gamma^{(m)}(t)$ 为 L 可积，$\{t: \gamma^{(k)}(t)\neq 0\}\subset\{t: \gamma(t)>0\}$，a. e. L，则 $g(\theta)$ 可估，其 UMVUE 为

$$\varphi(t)=\begin{cases}\gamma^{(k)}(t)/\gamma(t), & \gamma(t)>0;\\ 0, & \gamma(t)=0.\end{cases}\tag{4.21}$$

证 因为 $e^{-\theta t}\gamma^{(m)}(t)$ 为 L 可积，故存在 $a_n\to-\infty$、$b_n\to\infty$，使 $e^{-\theta a_n}\gamma^{(m)}(a_n)\to 0$，$e^{-\theta b_n}\gamma^{(m)}(b_n)\to 0\,(n\to\infty;\ m=0,\ \cdots,\ k)$. 应用分部积分法，得

$$\int_{a_n}^{b_n}e^{-\theta t}\gamma(t)dt=-\frac{1}{\theta}\gamma(t)e^{-\theta t}\Big|_{a_n}^{b_n}+\frac{1}{\theta}\int_{a_n}^{b_n}e^{-\theta t}\gamma^{(1)}(t)dt,$$

令 $n\to\infty$，得 $\displaystyle\int_{-\infty}^{\infty}e^{-\theta t}\gamma(t)dt=\frac{1}{\theta}\int_{-\infty}^{\infty}e^{-\theta t}\gamma^{(1)}(t)dt.$

用归纳法可证

$$\int_{-\infty}^{\infty}e^{-\theta t}\gamma(t)dt=\frac{1}{\theta^k}\int_{-\infty}^{\infty}e^{-\theta t}\gamma^{(k)}(t)dt.$$

但因为 $\displaystyle 1=c(\theta)\int_{-\infty}^{\infty}e^{-\theta t}\gamma(t)dt=\frac{c(\theta)}{\theta^k}\int_{-\infty}^{\infty}e^{-\theta t}\gamma^{(k)}(t)dt,$

所以有 $\displaystyle \theta^k=c(\theta)\int_{-\infty}^{\infty}e^{-\theta t}\gamma^{(k)}(t)dt.$

又因 $\{\gamma^{(k)}(t)\neq 0\}\subset\{\gamma(t)>0\}$，故(4.21)式所定义的 $\varphi(t)$ 为 θ^k 的无偏估计，再由 B-R-L-S 定理，本例即得证.

例 4.4 设 $X_1,\ \cdots,\ X_n$ 为 *iid.* 随机变量. X_1 具有密度

$$G(x_1;\alpha,\ P)=\begin{cases}\dfrac{\alpha^p}{\Gamma(P)}e^{-\alpha x_1}x_1^{p-1}, & \text{当 } 0<x_1\\ 0, & \text{当 } 0\leqslant x_1\end{cases}$$

$$(P>0\text{ 已知},0<\alpha<\infty).$$

现来估计 $g(\alpha)=\alpha^r$，r 为使 $np-r>0$ 的常数.

因为 $(X_1,\ \cdots,\ X_n)$ 的密度为

$$dP_\alpha(x_1, \cdots, x_n)$$
$$=\begin{cases}\dfrac{\alpha^{np}}{(\Gamma(P))^n}e^{-\alpha\sum\limits_1^n x_i}\prod\limits_1^n x_i^{p-1}dx_1\cdots dx_n, & x_i\geqslant 0(i=1, \cdots, n);\\ 0, & \text{其它}.\end{cases}$$

这是单参数指数族. 由定理 4.1 知, $T=\sum\limits_1^n x_i$ 是完全充分统计量. T 的密度为

$$G(t; \alpha, np)=\begin{cases}c(\alpha)e^{-\alpha t}\gamma(t)dt, & t>0;\\ 0, & t\leqslant 0.\end{cases}$$

其中 $c(\alpha)=\dfrac{\alpha^{np}}{\Gamma(np)}$, $\gamma(t)=t^{np-1}$.

在复平面区域$\{s: s=\alpha+iu, \alpha>0\}$内, $\dfrac{g(s)}{c(s)}=\dfrac{\Gamma(np)}{s^{np-r}}$解析,当 $np-r>0$ 时,存在拉氏逆变换,通过查表,可得其逆变换为

$$W(t)=\frac{\Gamma(np)}{\Gamma(np-r)}t^{np-r-1}, \quad t>0.$$

于是由定理 4.2 知, $g(\alpha)=\alpha^r$ 的 UMVUE 为

$$\hat{g}(t)=W(t)/\gamma(t)=\frac{\Gamma(np)}{\Gamma(np-r)}t^{-r}, \quad t>0.$$

如果 $np-2r>0$, 利用 Γ 积分容易算出此估计在平方损失下的风险为

$$\mathrm{Var}_\alpha\hat{g}(T(X))=\frac{\alpha^{np}}{\Gamma(np)}\int_0^\infty e^{-\alpha t}t^{np-1}\left(\frac{\Gamma(np)}{\Gamma(np-r)}\right)^2 t^{-2r}\,dt-\alpha^{2r}$$
$$=\left[\frac{\Gamma(np)\Gamma(np-2r)}{\Gamma(np-r)\Gamma(np-r)}-1\right]\alpha^{2r}.$$

当 r 为自然数时,利用例 4.3 中的公式(4.21),可以算得相同的结果.

例 4.5 设 $X_1, \cdots, X_n$ 为 *iid*. 随机变量, $X_1\sim N(\theta, 1)$ $(-\infty<\theta<\infty)$. 估计密度函数 $g_{x_0}(\theta)=\dfrac{1}{\sqrt{2\pi}}e^{-\frac{(x_0-\theta)^2}{2}}$, 其中 x_0 是一固定的实数; 假定 $n>1$.

由定理 4.1 知, $T=-n\bar{x}$ 为完全充分统计量. 因 $T\sim N(-n\theta, n)$, T 的密度为

$$dP_\theta^T(t)=c(\theta)e^{-\theta t}\gamma(t)dt,$$

这里$c(\theta)=\frac{1}{\sqrt{2n\pi}}e^{-\frac{n\theta^2}{2}}$，$\gamma(t)=e^{-\frac{t^2}{2n}}$. T 的特征函数为

$\phi_\theta(u)=\exp\{-in\theta u-nu^2/2\}$.

因为

$$\begin{aligned}&\phi_\theta(-u)g_{x_0}(\theta+iu)\\&=g_{x_0}(\theta)\exp\left\{i(x_0+(n-1)\theta)u-\frac{1}{2}(n-1)u^2\right\},\end{aligned}$$

于是注意$n>1$,有

$$\lim_{|u|\to\infty}|\phi_\theta(-u)g_{x_0}(\theta+iu)|=\lim_{|u|\to\infty}g_{x_0}(\theta)e^{-\frac{1}{2}(n-1)u^2}=0,$$

在$\theta\in[a,b]$中一致成立,$a<b$为任二实数.又有

$$\int_{-\infty}^{\infty}|\phi_\theta(-u)g_{x_0}(\theta+iu)|du=g_{x_0}(\theta)\int_{-\infty}^{\infty}e^{-\frac{1}{2}(n-1)u^2}du<\infty.$$

显然,$g_{x_0}(s)=\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}(x_0-s)^2\right\}$在全平面上解析,从而定理4.2的诸条件得到满足.$g_{x_0}(s)/c(s)$的双边拉氏逆变换为

$$\begin{aligned}w(t)&=\frac{1}{2\pi i}\int_{-i\infty}^{i\infty}\frac{g_{x_0}(s)}{c(s)}e^{ts}ds\\&=\frac{1}{2\pi i}\int_{-i\infty}^{i\infty}\left(\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x_0-s)^2}\right)\Big/\left(\frac{1}{\sqrt{2n\pi}}e^{-\frac{ns^2}{2}}\right)e^{ts}ds\\&=\sqrt{\frac{n}{n-1}}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x_0^2}e^{-\frac{1}{2}\frac{(x_0+t)^2}{n-1}}.\end{aligned}$$

利用公式(4.15)，$g_{x_0}(\theta)$的 UMVUE 为

$$\hat{g}_{x_0}(t)=\frac{w(t)}{\gamma(t)}=\sqrt{\frac{n}{n-1}}\cdot\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\frac{(nx_0+t)^2}{n(n-1)}\right\},$$

或者

$$\hat{\phi}_{x_0}(\bar{x})=\hat{g}_{x_0}(-n\bar{x})=\frac{1}{\sqrt{2\pi}\sqrt{1-\frac{1}{n}}}\exp\left\{-\frac{1}{2}\frac{(x_0-\bar{x})^2}{\left(1-\frac{1}{n}\right)}\right\}.$$

下面来计算$\hat{\phi}_{x_0}(\bar{X})$的方差，注意有一般结果：设$Y$、$W$是两

个随机变量，$W\sim N(\theta,\tau^2)$，$Y|W\sim N(W,\sigma^2)$，则 $Y\sim N(\theta,\tau^2+\sigma^2)$．在此令 $W=\bar{X}\left(\sim N\left(\theta,\frac{1}{n}\right)\right)$，$Y|\bar{X}\sim N\left(\bar{X},\frac{1}{2}\left(1-\frac{1}{n}\right)\right)$，那么 $Y\sim N\left(\theta,\frac{1}{n}+\frac{1}{2}\left(1-\frac{1}{n}\right)\right)=N\left(\theta,\frac{1}{2}\frac{1+n}{n}\right)$．于是有

$$
\begin{aligned}
f_Y(y;\theta)&=\int_{-\infty}^{\infty}f_{\bar{X}}(\bar{x};\theta)f_{Y|\bar{X}}(\bar{x},y;\theta)d\bar{x}\\
&=\frac{\sqrt{n}}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{n}{2}(\bar{x}-\theta)^2}\frac{1}{\sqrt{2\pi}\sqrt{\frac{1}{2}\left(1-\frac{1}{n}\right)}}\\
&\quad\times e^{-\frac{1}{2\left(\frac{1}{2}\left(1-\frac{1}{n}\right)\right)}(y-\bar{x})^2}d\bar{x}\\
&=\frac{1}{\sqrt{2\pi}\sqrt{\frac{1}{2}\frac{1+n}{n}}}e^{-\frac{(y-\theta)^2}{\left(\frac{1+n}{n}\right)}}.
\end{aligned}
\tag{4.22}
$$

于是

$$
\begin{aligned}
\mathrm{Var}_\theta\hat{\phi}_{x_0}(\bar{X})&=\frac{\sqrt{n}}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{n}{2}(\bar{x}-\theta)^2}\\
&\quad\times\left(\frac{1}{\sqrt{2\pi}\sqrt{1-\frac{1}{n}}}\exp\left(-\frac{1}{2}\frac{(x_0-\bar{x})^2}{1-\frac{1}{n}}\right)\right)^2d\bar{x}\\
&\quad-\left(\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(x_0-\theta)^2}{2}\right)\right)^2\\
&=\frac{1}{\sqrt{2\pi}\sqrt{2\left(1-\frac{1}{n}\right)}}\frac{\sqrt{n}}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{n}{2}(\bar{x}-\theta)^2}\\
&\quad\times\frac{1}{\sqrt{2\pi}\sqrt{\frac{1}{2}\left(1-\frac{1}{n}\right)}}e^{-\frac{1}{\left(1-\frac{1}{n}\right)}(x_0-\bar{x})^2}d\bar{x}\\
&\quad-\frac{1}{2\pi}\exp(-(x_0-\theta)^2)\\
&=\frac{1}{\sqrt{2\pi}\sqrt{2\left(1-\frac{1}{n}\right)}}\cdot\frac{1}{\sqrt{2\pi}\sqrt{\frac{1}{2}\frac{1+n}{n}}}
\end{aligned}
$$

$$\times e^{-\frac{(x_0-\theta)^2}{1+\frac{1}{n}}}-\frac{1}{2\pi}e^{-(x_0-\theta)^2}$$

$$=\frac{1}{2\pi}e^{-(x_0-\theta)^2}\left[\frac{n}{\sqrt{n^2-1}}e^{\frac{1}{1+n}(x_0-\theta)^2}-1\right].$$

其中第三步利用了(4.22)式，把 y 换成 x_0.

最后，若记

$$\hat{P}_\tau(\overline{X})=\int_{-\infty}^{\tau}\hat{\phi}_{x_0}(\overline{X})dx_0=\Phi\left(\frac{\tau-\overline{X}}{\sqrt{1-\frac{1}{n}}}\right),$$

利用 Fubini 定理，可得

$$E_\theta\hat{P}_\tau(\overline{X})=\int_{-\infty}^{\tau}E_\theta\hat{\phi}_{x_0}(\overline{X})dx_0=\int_{-\infty}^{\tau}\frac{1}{\sqrt{2\pi}}e^{-\frac{(x_0-\theta)^2}{2}}dx_0$$

$$=P_\theta[X_1<\tau]=\Phi(\tau-\theta).$$

可见 $\hat{P}_\tau(\overline{X})$ 为 $\Phi(\tau-\theta)$ 的 UMVUE.

§5 位置参数与刻度参数的无偏估计

一、截断型位置参数的无偏估计

根据 R. F. Tate[8] 的分类，位置参数可分为两种类型：截断型的(trancation)和转移型的(translation). 本段扼要介绍下面两种单边截断型位置参数族的无偏估计问题.

I 型：关于一维 L 测度的分布密度为

$$f(x;\theta)=K_1(\theta)h_1(x),\quad a<\theta<x<b. \tag{5.1}$$

其中 $h_1(x)$ 是 (a, b) 上定义的非负可测函数，对于任意的

$$\theta\in(a, b),\ 0<\int_\theta^b h_1(x)dx<\infty,\ K_1(\theta)=\left[\int_\theta^b h_1(x)dx\right]^{-1}.$$

II 型：关于一维 L 测度的分布密度为

$$f(x;\theta)=K_2(\theta)h_2(x),\quad a<x<\theta<b. \tag{5.2}$$

其中 $h_2(x)$ 是 (a, b) 上定义的非负可测函数，对于任意的

$$\theta\in(a, b),\ 0<\int_a^\theta h_2(x)dx<\infty,\ K_2(\theta)=\left[\int_a^\theta h_2(x)dx\right]^{-1}.$$

卢昆亮、赵林城还研究了双边截断型位置参数的无偏估计问

题，关于这方面的结果可参看[14].

首先，对上述定义的 I、II 作几点说明：(i)a、b 可为有限数，也可为 $-\infty$、∞；(ii)由定义可知，$K_1(\theta)$、$K_2(\theta)$在(a, b)的任一闭子区间上是绝对连续的，从而 $K_1'(\theta)$、$K_2'(\theta)$在(a, b)上 a.e.L 存在，并且对任意 $a<c<d<b(i=1, 2)$，有

$$K_i(d)-K_i(c)=\int_c^d K_i'(x)dx;$$

(iii)为了行文简便，以下均设 $h_i(x)>0$, $x\in(a, b)(i=1, 2)$.

引理 5.1 设 $X_1, \cdots, X_n$ 为 *iid.* 随机变量. 若 X_1 服从 I 型分布，则 $X_{(1)}=\min(X_1, \cdots, X_n)$为完全充分统计量；若 X_1 服从 II 型分布，则 $X_{(n)}=\max(X_1, \cdots, X_n)$为完全充分统计量.

证 仅证明 II 型情况：此时$(X_1, \cdots, X_n)$的密度为

$$\prod_{i=1}^n f(x_i, \theta)=(K_2(\theta))^n\prod_{i=1}^n h_2(x_i)I_{(a,\theta)}(x_{(n)}).$$

由因子判别定理立即可看出，$X_{(n)}$是充分统计量. 另一方面，$X_{(n)}$具有密度

$$f_{X(n)}(y; \theta)=\begin{cases}n(K_2(\theta))^n\left(\int_a^y h_2(x)dx\right)^{n-1}h_2(y), & a<y<\theta;\\ 0, & \text{其它}.\end{cases}$$

若对任一 $\theta\in(a, b)$, $E_\theta\phi(X_{(n)})=0$，则对任一 $\theta\in(a, b)$，有

$$\int_a^\theta\left[\int_a^y h_2(x)dx\right]^{n-1}h_2(y)\phi(y)dy=0.$$

由实变函数中的有关知识，有

$$\left[\int_a^y h_2(x)dx\right]^{n-1}h_2(y)\phi(y)=0, \text{ a. e. } L \text{ 于 } (a, b).$$

但对任一 $y\in(a, b)$有 $\int_a^y h_2(x)dx>0$, $h_2(y)>0$，故有

$$\phi(y)=0, \text{ a. e. } L \text{ 于}(a, b).$$

从而，对任一 $\theta\in(a, b)$，有

$$P_\theta[\phi(X_{(n)})\neq 0]$$
$$=n(K_2(\theta))^n\int_{\phi(y)\neq 0}\left(\int_0^y h_2(x)dx\right)^{n-1}h_2(y)dy=0,$$

这就证明了 $X_{(n)}$ 的完全性.

现在我们来考虑 I 型分布族的无偏估计问题: $g(\theta)$ 为待估实函数. 假定存在依赖于一个观察值 X_1 的无偏估计 $\phi(X_1)$, 则对任一 $\theta\in(a, b)$, 有

$$K_1(\theta)\int_\theta^b \phi(x)h_1(x)dx=g(\theta)$$

或

$$\int_\theta^b \phi(x)h_1(x)dx=\frac{g(\theta)}{K_1(\theta)}. \tag{5.3}$$

因上式左边及 $K_1(\theta)$ 在 (a, b) 的任一闭子区间上都是 θ 的绝对连续函数, 故 $g(\theta)$ 与 $\frac{g(\theta)}{K_1(\theta)}$ 亦然. 于是对(5.3)式两边求导后有

$$\phi(\theta)h_1(\theta)=-\left(\frac{g(\theta)}{K_1(\theta)}\right)', \quad \text{a. e. } L \text{ 于} (a, b), \tag{5.4}$$

再由(5.3)、(5.4)式可推出

$$\left(\frac{g(x)}{K_1(x)}\right)' \text{ 在} [c, b) \text{上可积(对任一} c\in(a, b)). \tag{5.5}$$

又因 $g(x)/K_1(x)$ 在 $[\theta, d]\subset(a, b)$ 上是绝对连续的, 故有

$$\begin{aligned}\frac{g(\theta)}{K_1(\theta)}&=\int_\theta^d -\left(\frac{g(x)}{K_1(x)}\right)'dx+\frac{g(d)}{K_1(d)}\\&=\int_\theta^d \phi(x)h_1(x)dx+\frac{g(d)}{K_1(d)}.\end{aligned} \tag{5.6}$$

令 $d\uparrow b$, 比较(5.6)式与(5.3)式, 可推得

$$\lim_{x\uparrow b}\frac{g(x)}{K_1(x)}=0. \tag{5.7}$$

可见(5.5)、(5.7)为 $g(\theta)$ 存在无偏估计 $\phi(X_1)$ 的必要条件. 反之, 假定(5.5)、(5.7)成立, 若令

$$\phi(X_1)=-\left(\frac{g(X_1)}{K_1(X_1)}\right)'\Big/h_1(X_1),$$

把上述过程倒推过去, 不难验证 $\phi(X_1)$ 确为 $g(\theta)$ 的无偏估计. 于是有如下定理:

定理 4.1 设 $X_1, \cdots, X_n$ 为 *iid.* 随机变量, X_1 具有 I 型的

密度．$g(\theta)$为在(a, b)上定义的实函数，那么它具有仅依赖于一个观察值的无偏估计的充要条件是(5.5)与(5.7)成立．当这些条件满足时，依赖于$X_1, \cdots, X_n$的$g(\theta)$的UMVUE存在且唯一，并为

$$\hat{g}(X_{(1)})=g(X_{(1)})-\frac{g'(X_{(1)})}{nK_1(X_{(1)})h_1(X_{(1)})}. \tag{5.8}$$

证 定理的前半部分以上已给出证明．并已得出：当(5.5)、(5.7)成立时，$\phi(X_1)=-\left(\dfrac{g(X_1)}{K_1(X_1)}\right)'\Big/h_1(X_1)$为$g(\theta)$的无偏估计．由引理5.1知，$X_{(1)}$为完全充分统计量，因此利用$B$-$R$-$L$-$S$定理，只需证明$(5.8)=E(\phi(X_1)\mid X_{(1)})$ a.s. 即可．

因为$X_1=X_{(1)}$的概率为$\dfrac{1}{n}$，而当$X_1>X_{(1)}$时，$X_{(1)}$即为$X_2, \cdots, X_n$的最小值，从而与X_1独立，所以当$x_1>x_{(1)}$时，有条件密度

$$\begin{aligned}
&f_{X_1\mid X_{(1)}}(x_1\mid x_{(1)})\\
&=\frac{K_1(\theta)h_1(x_1)(n-1)(K_1(\theta))^{n-1}h_1(x_{(1)})\left[\int_{x_{(1)}}^b h_1(t)dt\right]^{n-2}}{n(K_1(\theta))^n h_1(x_{(1)})\left[\int_{x_{(1)}}^b h_1(t)dt\right]^{n-1}}\\
&=\left(1-\frac{1}{n}\right)h_1(x_1)\Big/\int_{x_{(1)}}^b h_1(t)dt.
\end{aligned}$$

所以

$$E(\phi(X_1)\mid x_{(1)})=\frac{1}{n}\phi(x_{(1)})+\left(1-\frac{1}{n}\right)\frac{\int_{x_{(1)}}^b \phi(x_1)h_1(x_1)dx_1}{\int_{x_{(1)}}^b h_1(t)dt} \tag{5.9}$$

注意

$$\int_{x_{(1)}}^b \phi(x)h_1(x)dx=\frac{g(x_{(1)})}{K_1(x_{(1)})}; \tag{5.10}$$

$$\int_{x_{(1)}}^b h(t)\,dt=\frac{1}{K_1(x_{(1)})}. \tag{5.11}$$

对(5.11)式两边求导，可得

$$-\left(\frac{1}{K_1(x_{(1)})}\right)'\frac{1}{h_1(x_{(1)})}=1, \quad \text{a. e. } L \text{ 于}(a, b)$$

从而有

$$\begin{aligned}\phi(x_{(1)})&=-\left(\frac{g(x_{(1)})}{K_1(x_{(1)})}\right)'\Big/h(x_{(1)})\\&=-\left(\frac{1}{K_1(x_{(1)})}\right)'\frac{1}{h_1(x_{(1)})}g(x_{(1)})-\frac{g'(x_{(1)})}{K_1(x_{(1)})h_1(x_{(1)})}\\&=g(x_{(1)})-\frac{g'(x_{(1)})}{K_1(x_{(1)})h_1(x_{(1)})}, \quad \text{a. e. } L \text{ 于}(a, b). \qquad (5.12)\end{aligned}$$

把(5.10)、(5.11)、(5.12)代入(5.9)，便得(5.8).

注 若 $b<\infty$，当 $g(x)$在$[\theta, b]$上绝对连续时，(5.5)与(5.7)自然成立，这里 $\theta\in(a, b)$.

对于 II 型分布族，同样有

定理 5.2 设 $X_1, \cdots, X_n$ 为 *iid.* 随机变量，X_1 具有 II 型密度(5.2)，$g(\theta)$为在(a, b)上定义的实函数，则它具有仅依赖于一个观察值的无偏估计 $\phi(X_1)$的充要条件是

(i) $\left(\frac{g(x)}{K_2(x)}\right)'$ 在$(a, c]$上可积(对任一 $c\in(a, b)$)；

(ii) $\lim\limits_{x\downarrow a}\frac{g(x)}{K_2(x)}=0$.

在此二条件得到满足时，$g(\theta)$唯一地存在依赖于$(X_1, \cdots, X_n)$的 UMVUE：

$$\hat{g}(X_{(n)})=g(X_{(n)})+g'(X_{(n)})/(nK_2(X_{(n)})h_2(X_{(n)})). \qquad (5.13)$$

例 5.1 设 $X_1, \cdots, X_n$ 为 *iid.* 随机变量，$X_1\sim R(0, \theta)$，$0<\theta<\infty$. 因密度为

$$f(x; \theta)=\frac{1}{\theta} \quad (\text{当 } x\in(0, \theta)),$$

故此是 II 型截断族，这里 $h_2(x)=1$; $K_2(\theta)=\frac{1}{\theta}$.

令 $g_1(\theta)=\theta$，现在来估计 $g_1(\theta)$. 因它在$[0, c]$上绝对连续，故满足定理 5.2 的条件 (i) 与 (ii). 由公式(5.13)便得 θ 的 UMVUE 为

$$\hat{\theta}(X_{(n)})=\left(1+\frac{1}{n}\right)X_{(n)}.$$

同样可以得到 $\sqrt{\theta}$ 的 UMVUE 为

$$\hat{g}_2(X_{(n)})=\left(1+\frac{1}{2n}\right)\sqrt{X_{(n)}}.$$

而 $e^{\sqrt{\theta}}$ 的 UMVUE 为 $\hat{g}_3(X_{(n)})=\left(1+\frac{1}{2n}\sqrt{X_{(n)}}\right)e^{\sqrt{X(n)}}$.

X_1 的分布函数

$$F(z;\theta)=\begin{cases}0, & 当 z\leqslant 0\\ \frac{1}{\theta}z, & 当 0<z\leqslant\theta \quad (z 为已知实数)\\ 1, & 当 \theta<z\end{cases}$$

也满足定理 4.2 条件，其 UMVUE 为

$$\hat{g}_4(z;X_{(n)})=\begin{cases}0, & z\leqslant 0;\\ \left(1-\frac{1}{n}\right)\frac{z}{X_{(n)}}, & 0<z\leqslant X_{(n)};\\ 1, & X_{(n)}<z.\end{cases}$$

二、转移型位置参数的无偏估计

设 $f(x)$ 为一个一维分布密度. 我们称分布族

$$dP_\theta(x)=f(x-\theta)dx \quad (-\infty<\theta<\infty) \tag{5.14}$$

为带有转移型位置参数 θ 的分布族.

又设 $X_1, \cdots, X_n$ 为 *iid.* 随机变量，X_1 具有转移型密度 (5.14). 我们称统计量 $U(x_1, \cdots, x_n)$ 为转移型的，如果对任意的 $x_1, \cdots, x_n$ 及 c，有

$$U(x_1+c, \cdots, x_n+c)=U(x_1, \cdots, x_n)+c. \tag{5.15}$$

例如 $\overline{X}$, $X_{(1)}$, $X_{(n)}$ 等都是转移型的统计量. 对于转移型的统计量 U，如果有分布密度 $\varphi(u;\theta)$，则 θ 必定也为转移位置参数. 事实上，因为当 $X_i\sim f(x_i-\theta)$ 时，$X_i-\theta\sim f(x_i)$ $(i=1, \cdots, n)$. 故当 $U(X_1, \cdots, X_n)\sim\varphi(u;\theta)$ 时，$U(X_1-\theta, \cdots, X_n-\theta)\sim\varphi(u;0)$. 从而 $U(X_1, \cdots, X_n)=U(X_1-\theta, \cdots, X_n-\theta)+\theta\sim\varphi(u-\theta;0)$.

从这推得 $\varphi(u;\theta)=\varphi(u-\theta;0)$. 若记 $h(u)=\varphi(u;0)$，它是一密度，于是 $h(u-\theta)$ 即为参数为 θ 时 U 的密度，根据定义(5.14)，θ 为转移型位置参数.

设 $g(\theta)$ 为待估函数，以下我们来求解依赖于 U 的无偏估计：如果 $\hat{g}(U)$ 合乎所求，那么应有

$$\int_{-\infty}^{\infty}\hat{g}(u)h(u-\theta)du=g(\theta)\quad(-\infty<\theta<\infty).\tag{5.16}$$

令 $H(u)=h(-u)$，上式变为

$$\int_{-\infty}^{\infty}\hat{g}(u)H(\theta-u)du=g(\theta)\quad(-\infty<\theta<\infty).\tag{5.17}$$

即有 $\hat{g}*H=g$，g 为 $\hat{g}$ 与 H 的卷积. 因为 $H\in L_1(-\infty,\infty)$，如果 $\hat{g}\in L_1(-\infty,\infty)$，容易验证(5.17)式左边确有意义. 这样，当 g、H 满足适当条件时，我们便可以通过 Fourier 变换法或 Laplace 变换法由 H、g 解出 $\hat{g}$. 下面仅介绍 Fourier 变换求解法：

记 $\mathscr{F}(f;t)=\int_{-\infty}^{\infty}e^{itx}f(x)dx$ 为 f 的 Fourier 变换，对(4.17)式两边同时施加 Fourier 变换，因二函数卷积的 Fourier 变换等于此二函数的 Fourier 变换的乘积，从而有

$$\mathscr{F}(\hat{g};t)\mathscr{F}(H;t)=\mathscr{F}(g,t).$$

于是 $\mathscr{F}(\hat{g};t)=\mathscr{F}(g;t)/\mathscr{F}(H;t)$. 对此式两边同时进行 Fourier 逆变换，得

$$\hat{g}(u)=\mathscr{F}^{-1}(\mathscr{F}(g;t)/\mathscr{F}(H;t);u).\tag{5.18}$$

但是，上面的运算都是形式地进行的，要使每步能通过，应要求：

$$g\in L_1(-\infty,\infty);\tag{5.19}$$

$$\hat{g}\in L_1(-\infty,\infty);\tag{5.20}$$

$$\mathscr{F}(H,t)\neq 0\ (\text{对任一}\ t).\tag{5.21}$$

不过，如果由(5.18)式倒推回(5.16)式，只能保证(5.16)式在 a.e. L 意义下成立. 于是我们可得如下定理：

定理5.3 设 $X_1, \cdots, X_n$ 为 *iid.* 随机变量，X_1 具有密度(5.14)，U 为转移型统计量，具有密度 $h(u-\theta)$，并记 $H(u)=h(-u)$. 又假定(5.19)、(5.21)成立，则 $g(\theta)$ 如果存在一个 L_1 可积的无偏估计 $\hat{g}(u)$，那么必唯一地由(5.18)式给出. 反之，若 $g(\theta)$ 为连续函数，(5.19)、(5.21)式成立，$\mathscr{F}(g;t)/\mathscr{F}(H;t)$ 的 Fourier 逆变换存在，且其 L_1 可积，由(5.18)求出的 $\hat{g}(u)$ 使(5.16)式左端右(或左)连续，则 $\hat{g}$ 必为 $g(\theta)$ 的唯一的依赖于 U 的无偏估计. 又若 U 为完全充分统计量时，$\hat{g}(u)$ 是 $g(\theta)$ 的 UMVUE.

例5.2 设 $X_1, \cdots, X_n$ 为 *iid.* 随机变量，$X_1 \sim N(\theta, 1)$ $(-\infty<\theta<\infty)$，估计函数

$$g(\theta; x_0)=\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(x_0-\theta)^2}{2}\right\}.$$

显然，X_1 具有转移参数密度，$U=\bar{X}$ 为转移型统计量，且完全充分. 此时，U 在 $\theta=0$ 时的密度为

$$h(u)=\frac{\sqrt{n}}{\sqrt{2\pi}} e^{-\frac{n}{2}u^2}, \quad H(u)=h(-u)=h(u).$$

因为

$$\mathscr{F}(H;t)=e^{-\frac{t^2}{2n}}, \quad \mathscr{F}(g;t)=\exp\left\{ix_0 t-\frac{t^2}{2}\right\}.$$

故 $g(\theta; x_0)$ 的 UMVUE 为

$$\begin{aligned}
\hat{g}(\bar{x}; x_0) &= \mathscr{F}^{-1}(\mathscr{F}(g;t)/\mathscr{F}(H;t); \bar{x}) \\
&= \mathscr{F}^{-1}\left(\exp\left\{ix_0 t-\frac{1}{2}\left(1-\frac{1}{n}\right)t^2\right\}; \bar{x}\right) \\
&= \frac{1}{\sqrt{2\pi}\sqrt{1-\frac{1}{n}}} \exp\left(-\frac{(x_0-\bar{x})^2}{2\left(1-\frac{1}{n}\right)}\right).
\end{aligned}$$

此结果与例4.5是一致的.

三、刻度参数的无偏估计

1. 刻度参数族与齐次统计量　设随机变量 X 具有密度

$$dP_\rho(x)=f(x;\rho)dx=\rho f(\rho x)\quad(0<\rho<\infty),\qquad(5.22)$$

其中 $f(x)$ 为一个一维分布密度. 称此分布族为刻度参数族, ρ 为刻度参数(Scale parameter). 例如 $X\sim N(0,\sigma^2)$, $\rho=\frac{1}{\sigma}$ 就是刻度参数,此处

$$f(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2},\quad f\left(x;\frac{1}{\sigma}\right)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{x^2}{2\sigma^2}}.$$

又如 $X\sim R(0,\theta)$ $(0<\theta<\infty)$, $\frac{1}{\theta}$ 也是刻度参数,这里

$$f(x)=\begin{cases}1, & \text{当 } x\in[0,1],\\ 0, & \text{当其它;}\end{cases}$$

$$f\left(x;\frac{1}{\theta}\right)=\begin{cases}\frac{1}{\theta}, & \text{当 } x\in[0,\theta],\\ 0, & \text{其它.}\end{cases}$$

设 $X_1,\cdots,X_n$ 为 *iid*. 随机变量, X_1 具有刻度参数族密度(5.22). 我们称 $Y=H_\alpha(X_1,\cdots,X_n)$ 为幂次是 $\alpha(\alpha\neq0)$ 的齐次统计量, 如果对任意的 $\rho>0,x_1,\cdots,x_n$, 有 $H_\alpha(\rho x_1,\cdots,\rho x_n)=\rho^\alpha H_\alpha(x_1,\cdots,x_n)$.

如果幂次为 α 的齐次统计量 Y 存在密度, 则必具有形式

$$h(y;\rho)=\rho^\alpha h(\rho^\alpha y).\qquad(5.23)$$

事实上,当参数为 ρ 时,设 Y 具有密度 $h(y;\rho)$,但因为 $\rho X_i\sim f(x)$ $(i=1,\cdots,n)$,故 $H_\alpha(\rho X_1,\cdots,\rho X_n)\sim h(y;1)$. 从而 $Y=H_\alpha(X_1,\cdots,X_n)=\rho^{-\alpha}H_\alpha(\rho X_1,\cdots,\rho X_n)\sim\rho^\alpha h(\rho^\alpha y;1)$. 记 $h(y)=h(y;1)$, 于是(5.23)式得证.

定理 5.4 设 $X_1,\cdots,X_n$ 为 *iid*. 随机变量,具有分布(5.22), $Y=H_\alpha(X_1,\cdots,X_n)$ 为幂次是 $\alpha(\neq0)$ 的齐次统计量, 具有密度(5.23). 当 $0<E_1[Y^{-\frac{r}{\alpha}}]<\infty$ 时,则 $g(\rho)=\rho^r$ 可估,下式为其一无偏估计

$$\varphi(Y)=Y^{-\frac{r}{\alpha}}/E_1(Y^{-\frac{r}{\alpha}}).\qquad(5.24)$$

证 $E_\rho\varphi(Y)=\int_{-\infty}^{\infty}\varphi(y)\rho^\alpha h(\rho^\alpha y)dy$

$$=\int_{-\infty}^{\infty}\varphi(\rho^{-\alpha}z)h(z)dz \quad (\text{令 } \rho^\alpha y=z)$$

$$=\frac{1}{E_1(Y^{-r/\alpha})}\int_{-\infty}^{\infty}(\rho^{-\alpha})^{-r/\alpha}z^{-r/\alpha}h(z)dz$$

$$=\frac{\rho^r}{E_1(Y^{-\frac{r}{\alpha}})}E_1(Y^{-\frac{r}{\alpha}})=\rho^r.$$

例 5.3 设 $X_1, \cdots, X_n$ 为 *iid*. 随机变量，X_1 遵从 Weibull 分布：

$$f(x_1;\alpha,\rho)=\begin{cases}\alpha\rho(\rho x_1)^{\alpha-1}\exp(-(\rho x_1)^\alpha), & x_1>0;\\ 0, & \text{其它}\end{cases} \tag{5.25}$$

($\alpha>0$ 已知, $0<\rho<\infty$). 估计 $g(\rho)=\rho^r$.

$Y=\sum_1^n X_i^\alpha$ 是幂次为 α 的齐次统计量. 利用因子判别定理知 Y 是充分的. 若令 $\theta=\rho^\alpha$, $Z_i=X_i^\alpha$, 于是 Z_i 具有密度 $G(z;\theta,1)$. 故 $Y=\sum_1^n Z_i$ 具有密度 $G(y;\theta,n)$. 从而可知 Y 又是完全的. 当 $n>r/\alpha$ 时,

$$E_1(Y^{-\frac{r}{\alpha}})=\int_0^\infty y^{-\frac{r}{\alpha}}G(y;1,n)dy$$

$$=\frac{1}{\Gamma(n)}\int_0^\infty y^{-\frac{r}{\alpha}}y^{n-1}e^{-y}dy=\Gamma\left(n-\frac{r}{\alpha}\right)\Big/\Gamma(n).$$

由公式(5.24), 得 ρ^r 的 UMVUE 为

$$\varphi(Y)=Y^{-\frac{r}{\alpha}}\Gamma(n)/\Gamma\left(n-\frac{r}{\alpha}\right).$$

2. Mellin 变换及其在无偏估计中的应用　假若复变函数

$$M(f;s)=\int_0^\infty x^{s-1}f(x)dx \quad (s\text{ 为复变数}) \tag{5.26}$$

在某区域有定义, 则称其为函数 f 的 Mellin 变换. 若令 $x=e^{-y}$, 于是

$$M(f;s)=\int_{-\infty}^{\infty}e^{-ys}f(e^{-y})dy.$$

可见$f(x)$的 Mellin 变换即为$f(e^{-y})$的双边拉氏变换. 因此，相应于双边拉氏变换, Mellin 变换也具有如下性质:

(a) Mellin 变换的存在域为条形区域$\{s: a<\mathrm{Re}(s)<b\}$, $a<b$ 可为有限或为无限数，边界可能属于存在域， 也可能不属于存在域.

(b) 反演公式

$$f(x)=\frac{1}{2\pi i}\lim_{T\to\infty}\int_{c-iT}^{c+iT}M(f;s)x^{-s}\,ds,\ \text{a. e. } L,\ \ c\in(a,\ b). \tag{5.27}$$

(c) **表示定理**: 若$\Phi(s)$为在$\{s: a<\mathrm{Re}(s)<b\}$内解析, 并且(i)对任何 $[c,\ d]\subset(a,\ b)$, 有$\lim\limits_{s\to\infty}\Phi(s)=0$, 在$\{s:\ c\leqslant\mathrm{Re}(s)\leqslant d\}$中一致成立; (ii)对某个$c\in(a,\ b)$, 使

$$\int_{-\infty}^{\infty}|\Phi(c+it)|\,dt<\infty.$$

则存在函数$f(x)$, 使$M(f;s)$存在域包含$\{s: a<\mathrm{Re}(s)<b\}$, 并且在此区域内有$\Phi(s)=M(f;s)$.

(d) **褶积定理**. 设f、g在$(0,\ \infty)$上定义. 若积分

$$h(\rho)=\int_0^{\infty}f(x)g\left(\frac{\rho}{x}\right)\frac{dx}{x}\quad(\rho>0) \tag{5.28}$$

有意义, 则称函数$h(\rho)$为f与g的褶积, 记作$f\circledast g=h$. 容易验证有$f\circledast g=g\circledast f$.

假若$M(f;s)$的存在域与$M(g;s)$的存在域有公共部分$A=\{s: c<\mathrm{Re}(s)<d\}$, 则$h(\rho)=f\circledast g$在$\rho>0$上 a. e. L 存在且为有限的, 而且h也有 Mellin 变换$M(h;s)$, 其存在域包含A.

证 取实数$\sigma\in(c,\ d)$, 考察二元函数

$$K(x,\ \rho)=\rho^{\sigma-1}|f(x)|\,|g(\rho/x)|\frac{1}{x},\quad x>0,\ \rho>0.$$

$$\int_0^{\infty}\int_0^{\infty}K(x,\ \rho)\,dx\,d\rho=\int_0^{\infty}\frac{|f(x)|}{x}\left[\int_0^{\infty}\left|g\left(\frac{\rho}{x}\right)\right|\rho^{\sigma-1}\,d\rho\right]dx$$

$$=\int_0^{\infty}x^{\sigma-1}|f(x)|\,dx\int_0^{\infty}\rho_1^{\sigma-1}|g(\rho_1)|\,d\rho_1<\infty$$

$$(\text{令 }\rho_1=\rho/x),$$

那么,根据 Fubini 定理,$K(x, \rho)$的 a.e. $L.\ \rho$ 截口可积,即

$$\int_0^{\infty}|f(x)|\left|g\left(\frac{\rho}{x}\right)\right|\frac{dx}{x}<\infty, \quad \text{a.e. } L \text{ 于 } (0, \infty).$$

同时,还证明了 $M\{h; s\}=M\{f \circledast g; s\}=M(f; s)\cdot M(g; s)$.

设 $X_1, \cdots, X_n$ 为 *iid.* 随机变量,X_1 具有密度(5.22),Y 为一幂次为 $\alpha(\neq 0)$的齐次统计量,具有密度(5.23). $g(\rho)$为待估函数. 假若 $\hat{g}(Y)$为 $g(\rho)$的无偏估计,应有

$$\int_{-\infty}^{\infty}\hat{g}(y)\rho^{\alpha}h(\rho^{\alpha}y)dy=g(\rho), \ \rho\in(0, \infty), \qquad (5.29)$$

令 $\rho_1=\rho^{\alpha}$,并作变换 $z=1/y$,可得

$$\int_0^{\infty}\frac{1}{z}\hat{g}\left(\frac{1}{z}\right)h\left(\frac{\rho_1}{z}\right)\frac{dz}{z}+\int_0^{\infty}\frac{1}{z}\hat{g}\left(-\frac{1}{z}\right)h\left(-\frac{\rho_1}{z}\right)\frac{dz}{z}$$

$$=\frac{1}{\rho_1}g(\rho_1^{\frac{1}{\alpha}}), \quad \rho_1\in(0, \infty).$$

如果当 $x<0$ 时 $h(x)=0$,那么有

$$\int_0^{\infty}\frac{1}{z}\hat{g}\left(\frac{1}{z}\right)h\left(\frac{\rho_1}{z}\right)\frac{dz}{z}=\frac{1}{\rho_1}g(\rho_1^{\frac{1}{\alpha}}), \quad \rho_1\in(0, \infty),$$

即有

$$\frac{1}{z}\hat{g}\left(\frac{1}{z}\right)\circledast h(z)=\frac{1}{\rho}g(\rho^{\frac{1}{\alpha}}), \quad \rho\in(0, \infty). \qquad (5.30)$$

对上式两边进行 Mellin 变换,有

$$M\left\{\frac{1}{z}\hat{g}\left(\frac{1}{z}\right); s\right\}M\{h(z); s\}=M\left\{\frac{1}{z}g(z^{\frac{1}{\alpha}}); s\right\}.$$

从而可解出

$$\hat{g}\left(\frac{1}{z}\right)=z M^{-1}\left\{M\left(\frac{1}{z}g(z^{\frac{1}{\alpha}}); s\right)\Big/M(h; s); z\right\}$$

或

$$\hat{g}(y)=\left(\frac{1}{y}\right)M^{-1}\left\{M\left(\frac{1}{z}g(z^{\frac{1}{\alpha}}); s\right)\Big/M(h; s), \frac{1}{y}\right\}. \qquad (5.31)$$

把如上过程反推回去:若令 $\hat{g}(Y)$由(5.31)定义,则有

$$E_{\rho}(\hat{g}(Y))=g(\rho), \ \text{a.e.} L \text{于 } (0, \infty). \qquad (5.32)$$

假定 $g(\rho)$为连续函数,而上式左边为 ρ 的左(或右)连续函数,则由(5.32)可推出(5.29)式成立.

当然，上面的计算过程也是形式地进行的．但加上适当条件，这个计算过程是正确的．因此，我们有如下结果：

定理 5.5 在以上记号下，假定 $x<0$ 时 $h(x)=0$，(i)若 $g(\rho)$ 的无偏估计 $\hat{g}(Y)$ 存在，$\frac{1}{x}\hat{g}\left(\frac{1}{x}\right)$、$h(x)$ 的 Mellin 变换都存在，两个存在域有公共部分 $A=\{s: c<\mathrm{Re}(s)<d\}$，在 A 上 $M(h;s)\neq 0$，则 $g(\rho)$ 的基于 Y 的无偏估计 $\hat{g}(Y)$ 是唯一的，由(5.31)式决定．(ii)又设在某带形区域内，函数

$$K(s)=M\left[\frac{1}{z}g(z^{\frac{1}{\lambda}});s\right]\Big/M(h;s)$$

满足表示定理条件，并且 $g(\rho)$ 为连续函数，由(5.31)式决定的 $\hat{g}(y)$ 使 $E_\rho[\hat{g}(Y)]$ 对 ρ 左(或右)连续，则 $\hat{g}(Y)$ 是 $g(\rho)$ 的一个无偏估计．

此定理的具体应用可参看[13]，p. 147．

以上分别讨论了位置参数和刻度参数的无偏估计．至于位置参数与刻度参数同时估计的情况，可参看[13]等．

§6 线性模型中参数无偏估计

一、矩阵的广义逆与最小二乘法

在线性模型的研究中需要一些矩阵的广义逆的基本知识，这里简要地作一些介绍，关于这方面的进一步内容可参看 C. R. Rao[10] 的第一章．

设 A 是一 $m\times n$ 阶矩阵，若存在一个 $n\times m$ 阶矩阵 B，使 $ABA=A$，则称 B 为 A 的一个广义逆．常以 A^- 表示 A 的广义逆．显然，若 A 为满秩方阵时，A^{-1} 必为 A 的广义逆，且 A 仅有这个广义逆．因此，广义逆可看成是对通常逆矩阵的推广．广义逆有如下一些性质：假定 A 为 $m\times n$ 阶矩阵．

(i) 对任意矩阵 A，A^- 总是存在的．

证 设 A 的秩为 r，于是存在满秩的 m 阶和 n 阶方阵

P、Q, 使 $A=P\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}Q$, 其中 I_r 表示 r 阶单位阵. 记 $A^-=Q^{-1}\begin{bmatrix} I_r & 0 \\ 0 & L \end{bmatrix}P^{-1}$, 其中 L 为任意的 $(n-r)\times(m-r)$ 阶阵, 不难验证有 $AA^-A=A$, 即 A^- 确为 A 的广义逆. 特别取 $L=0$ 时, 有 $R(A^-)=R(A)$. 当 $(n-r)(m-r)\neq 0$ 时, 由 L 的任意性知, A 具有无穷多个广义逆. 还可进一步证明, 当 $r<\max\{m, n\}$ 时, A 有无穷多的广义逆.

(ii) A^- 为 A 的广义逆⇔对任一相容方程 $AX=Y$, A^-Y 为其解.

证 "⇒"因为当 $AX=Y$ 相容时, 则有解 X_0, 使 $AX_0=Y$, 于是 $A(A^-Y)=AA^-AX_0=AX_0=Y$, 即 A^-Y 为解.

"⇐"取 $Y=A_i$(A 的第 i 列), 则方程 $AX=A_i$ 相容. 因 A^-A_i 为解, 则有 $AA^-A_i=A_i$, 令 $i=1, \cdots, n$, 从而有

$$AA^-A=AA^-(A_1\vdots\cdots\vdots A_n)=(A_1\vdots\cdots\vdots A_n)=A.$$

(iii) $H=A^-A$ 为幂等阵, 且 $R(H)=\mathrm{trace}(H)=R(A)$.

证 因为 $HH=A^-AA^-A=A^-A=H$, 故 H 幂等. 由幂等阵的性质知, $R(H)=\mathrm{trace}(H)$. 又由于 $H=A^-A$, 故 $R(H)\leqslant R(A)$, 但另一方面又由于 $A=AH$, 故又有 $R(A)\leqslant R(H)$, 从而得 $R(H)=R(A)$.

(iv) 相容方程 $AX=Y$ 的通解为

$$X=A^-Y+(A^-A-I)Z, \tag{6.1}$$

其中 A^- 为 A 某一固定的广义逆, Z 为任意 n 维向量.

证 容易验证(6.1)确为方程的解. 反之, 若 X_0 为任一解, 那么

$$X_0=A^-AX_0+(A^-A-I)(-X_0)=A^-Y+(A^-A-I)(-X_0)$$

即为(6.1)式的形式.

(v) C 为任一给定的 n 维向量, 则对相容方程 $AX=Y$ 的任一解 X, 使 $C'X$ 唯一 ⇔ $C'A^-A=C'$. 这里及以下的撇号("$'$")表示转置.

证 由(6.1)式知，$C'X$ 唯一$\Leftrightarrow C'A^-Y+C'(A^-A-I)Z$ 唯一，Z 任意$\Leftrightarrow C'A^-A=C'$.

(vi) $(A^-)'$ 为 A' 的广义逆.

证 因为 $A'(A^-)'A'=(AA^-A)'=A'$.

(vii) 对 $A'A$ 的任一广义逆 $(A'A)^-$，有

(a) $A'A(A'A)^-A'=A',\ A(A'A)^-AA'=A$;

(b) $A(A'A)^-A'$ 是幂等的、不变的.

证 (a)因为

$$\begin{aligned}&(A'A(A'A)^-A'-A')(A'A(A'A)^-A'-A')'\\&=A'A(A'A)^-A'A(A'A)^-A'A-A'A(A'A)^-A'A\\&\quad-A'A(A'A)^-A'A+A'A=0,\end{aligned}$$

故(a)的前一等式得证，后一等式可同样证明.

(b) $A(A'A)^-A'$ 的幂等性由(a)可直接验证. 为证其不变性，利用线性代数中一个事实：$\mu(A')=\mu(A'A)$，此处及以后的记号 $\mu(A)$ 表示以矩阵 A 的列向量张成的线性空间. 于是存在矩阵 C，使 $A'=A'AC$，从而 $A(A'A)^-A'=C'A'A(A'A)^-A'AC=C'A'AC$. 这证明了 $A(A'A)^-A'$ 的值与 $(A'A)^-$ 的取法无关.

下面转向最小二乘法：设 Y 是已知的 n 维向量，X 为已知的 $n\times p$ 阶常数矩阵，β 为未知的 p 维向量. 我们有如下熟知的结果：

定理 6.1 p 维向量 $\hat{\beta}$ 使 $(Y-X\beta)'(Y-X\beta)$ 的值达到最小的充要条件是：$\hat{\beta}$ 满足方程

$$X'X\beta=X'Y. \tag{6.2}$$

此方程总是相容的，且

$$\hat{\beta}=(X'X)^-X'Y \tag{6.3}$$

为解.

以后我们称方程(6.2)为正规方程，其任一解 $\hat{\beta}$ 称为最小二乘解.

证 因为

$$\frac{\partial}{\partial\beta}[(Y-X\beta)'(Y-X\beta)]$$

$$=\frac{\partial}{\partial\beta}[Y'Y-2Y'X\beta+\beta'X'X\beta]=2(X'X\beta-X'Y),$$

令其 $=0$, 便推得正规方程(6.2). 这证明了使 $(Y-X\beta)'(Y-X\beta)$ 达到最小的 $\hat{\beta}$ 必然满足(6.2).

因为 $\mu(X'Y)\subset\mu(X')=\mu(X'X)$, 故方程(6.2)是相容的, 又根据广义逆的性质(ii): (6.3)为(6.2)的解.

假定 $\hat{\beta}$ 是(6.2)的任一解, β 为任一 p 维向量. 注意

$$(X\hat{\beta}-X\beta)'(Y-X\hat{\beta})=(\hat{\beta}-\beta)(X'Y-X'X\hat{\beta})=0,$$

从而有

$$\begin{aligned}&(Y-X\beta)'(Y-X\beta)\\&\quad=(Y-X\hat{\beta}+X\hat{\beta}-X\beta)'(Y-X\hat{\beta}+X\hat{\beta}-X\beta)\\&\quad=(Y-X\hat{\beta})'(Y-X\hat{\beta})+(\hat{\beta}-\beta)'X'X(\hat{\beta}-\beta)+0\\&\quad\geqslant(Y-X\hat{\beta})'(Y-X\hat{\beta})\end{aligned}$$

于是定理得证.

需要注意的是: 形如(6.3)的最小二乘解, 未必构成最小二乘解的全体, 因由广义逆的性质(iv)知, 方程(6.2)的解的全体为

$$\hat{\beta}=(X'X)^{-}X'Y+((X'X)^{-}(X'X)-I)Z, \qquad (6.4)$$

其中 Z 为任意 p 维向量. 不过, 以后我们将会看到, 在线性模型中, 只考虑形如(6.3)的最小二乘解就够了.

二、Gauss-Markov 模型

考虑线性模型

$$y_i=x_{i1}\beta_1+\cdots+x_{ip}\beta_p+e_i \quad (i=1,\ \cdots,\ n) \qquad (6.5)$$

或表示为 $Y=X\beta+e$, 其中 $X=(x_{ij})_{n\times p}$ 为已知常数矩阵, 称为设计矩阵, $Y=(y_1,\ \cdots,\ y_n)'$ 为可观察的随机向量, $e=(e_1,\ \cdots,\ e_n)'$ 为随机误差向量, 不能量测, 而 $\beta=(\beta_1,\ \cdots,\ \beta_p)'$ 为未知的回归系数.

假设随机误差向量 e 服从参数为 $(\sigma^2,\ \varphi)$ 的分布 $P(\sigma^2,\ \varphi)$,

$0<\sigma<\infty$，并满足条件

$$\begin{cases} E_{(\sigma^2,\varphi)}e=0, \\ E_{(\sigma^2,\varphi)}(e_ie_j)=\sigma^2\delta_{ij} \quad (i,\ j=1,\ \cdots,\ n). \end{cases} \tag{6.6}$$

则称此时的线性模型(6.5)为 Gauss-Markov 模型. 显然，在此模型下，Y 的分布依赖于参数$(\beta,\ \sigma^2,\ \varphi)$，其中 $0<\sigma<\infty$，β 为 p 维向量，φ 是一般抽象参数. 由(6.6)立得

$$E_{(\beta,\sigma^2,\varphi)}(Y)=X\beta; \tag{6.7}$$

$$\mathrm{Var}_{(\beta,\sigma^2,\varphi)}(Y)=\sigma^2 I_n. \tag{6.8}$$

以后，在含义明确时，(6.7)、(6.8)式左边中的参数$(\beta,\ \sigma^2,\ \varphi)$略而不写. 值得注意的是，在上述模型里，分布族中既有含意明确的向量参数$(\beta,\ \sigma^2)$，又含有一般性的参数 φ，对 φ 的要求是使(6.6)式成立. 因此 Gauss-Markov 模型既带有参数性质，又带有非参数性质.

设 C 为某一 p 维向量，如果存在 n 维向量 L，使 Y 的线性函数 $L'Y$ 满足

$$E(L'Y)=C'\beta \quad (\text{对任意 } \beta), \tag{6.9}$$

则称线性函数 $C'\beta$ 是线性可估的，$L'Y$ 为 $C'\beta$ 的线性无偏估计.

容易验证，$L'Y$ 为 $C'\beta$ 的无偏估计的充要条件是 $L'X=C'$. 事实上，若对任意 β，有 $E(L'Y)=L'E(Y)=L'X\beta=C'\beta$，必可推出 $L'X=C'$. 反之，显然也成立.

假定 $C'\beta$ 为线性可估的，记 $\mathscr{L}=\{L: L'X=C'\}$. 若有 $L_0\in\mathscr{L}$，使

$$\mathrm{Var}(L_0'Y)=\min_{L\in\mathscr{L}}\mathrm{Var}(L'Y), \tag{6.10}$$

则称 $L_0'Y$ 为 $C'\beta$ 的最优线性无偏估计(The best linear unbiased estimate)，简记为 BLUE.

由(6.8)式，我们有 $\mathrm{Var}(L'Y)=\sigma^2L'L$. 故(6.10)等价于

$$L_0'L_0=\min_{L\in\mathscr{L}}L'L. \tag{6.11}$$

下面，我们来讨论 $C'\beta$ 线性可估的充要条件以及求解 BLUE 的方法.

定理 6.2 设 C 是一个 p 维向量，在 Gauss-Markov 模型里，

以下诸条件等价：

(i) $C'\beta$ 可估；(ii) $C'\beta$ 线性可估；(iii) $C\in\mu(X'X)$ 或 $R(X')=R(X' \vdots C)$；(iv) 对任一最小二乘解 $\hat{\beta}$，$C'\hat{\beta}$ 是唯一的；(v) $C'=C'(X'X)^{-}X'X$；(vi) $C'\hat{\beta}$ 是 $C'\beta$ 的无偏估计，其中 $\hat{\beta}$ 为任一最小二乘解.

证 (ii)$\Rightarrow$(iii). 设存在 $C'\beta$ 的线性无偏估计 $L'Y$，由前所述，有 $C'=L'X$，即 $C=X'L$，从而 $C\in\mu(X')=\mu(X'X)$.

(iii)$\Rightarrow$(iv). 由(iii)可设 $C=X'XD$，其中 D 是某一向量，利用(6.4)式，并注意广义逆的性质(vii)中(a)、(b)，有

$$
\begin{aligned}
C'\hat{\beta} &= D'X'X(X'X)^{-}X'Y + D'X'X((X'X)^{-}X'X-I)Z \\
&= D'X'Y.
\end{aligned}
$$

可见 $C'\hat{\beta}$ 与 $\hat{\beta}$ 的选择无关.

(iv)$\Leftrightarrow$(v)，由(6.4)式及广义逆的性质(v)即得.

(v)$\Rightarrow$(vi)，由(iv)知 $C'\hat{\beta}$ 唯一，故 $\hat{\beta}$ 可取(6.3)式所表达的最小二乘解，又根据(v)：

$$
\begin{aligned}
EC'\hat{\beta} &= E[C'(X'X)^{-}X'Y] \\
&= E[C'(X'X)^{-}(X'X)(X'X)^{-}X'Y] \\
&= C'(X'X)^{-}(X'X)(X'X)^{-}X'X\beta \\
&= C'(X'X)^{-}X'X\beta = C'\beta.
\end{aligned}
$$

(vi)$\Rightarrow$(ii)，因为(6.3)式为 Y 的线性函数，故 $C'\hat{\beta}$ 为 $C'\beta$ 的线性无偏估计.

最后，显然有(ii)$\Rightarrow$(i)，所以如果能证明(i)$\Rightarrow$(iii)，就完成了整个定理的证明. 设 $\hat{g}(Y)$ 为 $C'\beta$ 的无偏估计，则对任意的 β，有 $E_{(\beta,\sigma^2,\varphi)}\hat{g}(Y)=C'\beta$. 注意：当参数为 $(\beta, \sigma^2, \varphi)$ 时，$Y-X\beta$ 的分布与当参数为 $(0, \sigma^2, \varphi)$ 时 Y 的分布相同，故对任意的使 $X\beta=0$ 的 β，有

$$
\begin{aligned}
C'\beta &= E_{(\beta,\sigma^2,\varphi)}\hat{g}(Y) = E_{(\beta,\sigma^2,\varphi)}\hat{g}(Y-X\beta) \\
&= E_{(0,\sigma^2,\varphi)}\hat{g}(Y) = C'0 = 0.
\end{aligned}
$$

即推得 $\mu(X')^{\perp}\subset\mu(C)^{\perp}$，从而有 $C\in\mu(X')$，(iii)成立. 定理证毕.

此定理告诉我们：β 的线性函数 $C'\beta$ 的可估性与线性可估性是等价的．因此，以后我们对其可估与线性可估不加区别．定理中所列 $C'\beta$ 可估的诸充要条件中，第(iii)条是最常用的条件．当 $R(X')=p$，即 $X'X$ 满秩时，因对任意的 p 维向量 C，总有 $C\in\mu(X'X)$，故由(iii)推得此时 β 的任意线性函数均是可估的，特别是，若 β 的每一分量可估，称 β 为可估的，$\hat{\beta}=(X'X)^{-1}X'Y$ 为 β 的唯一的线性无偏估计．

当 $C'\beta$ 可估时，$C'\hat{\beta}$ 称为 $C\beta$ 的最小二乘估计(least squares estimate)，简记为 LS 估计，其中 $\hat{\beta}$ 为正规方程 $X'X\beta=X'Y$ 的任一解，特别是，可取 $\hat{\beta}=(X'X)^{-}X'Y$．

定理 6.3 在 Gauss-Markov 模型里，可估函数 $C'\beta$ 的 LS 估计 $C'\hat{\beta}$ 是 $C'\beta$ 的唯一的 BLUE．设 $D'\beta$ 是另一可估线性函数，则有

$$\operatorname{cov}(C'\hat{\beta},\ D'\hat{\beta})=C'(X'X)^{-}D\sigma^2. \tag{6.12}$$

特别是，当 $R(X)=p$ 时，有

$$\operatorname{Var}(\hat{\beta})=(X'X)^{-1}\sigma^2. \tag{6.13}$$

证 因 $C'\beta$ 可估，那么对任一最小二乘解 $\hat{\beta}$，$C'\hat{\beta}$ 是唯一的，因此以下证明中可取 $\hat{\beta}=(X'X)^{-}X'Y$．由定理 6.2.(v)，我们有

$$C'=C'(X'X)^{-}X'X. \tag{6.14}$$

设 $L'Y$ 是 $C'\beta$ 的任一线性无偏估计，则有 $C'=L'X$．于是由上式及(6.14)，有

$$\begin{aligned}&(L'-C'(X'X)^{-}X')X(X'X)^{-}C\\&\quad=(L'X-C'(X'X)^{-}X'X)(X'X)^{-}C\\&\quad=(C'-C')(X'X)^{-}C=\mathbf{0}.\end{aligned}$$

从而

$$\begin{aligned}\operatorname{Var}(L'Y)&=L'L\sigma^2\\&=(L-X(X'X)^{-}C+X(X'X)^{-}C)'\\&\quad\times(L-X(X'X)^{-}C+X(X'X)^{-}C)\sigma^2\\&=(L-X(X'X)^{-}C)'(L-X(X'X)^{-}C)\sigma^2\\&\quad+C'(X'X)^{-}X'X(X'X)^{-}C\sigma^2\\&\geqslant(C'(X'X)^{-}X')(X(X'X)^{-}C)\sigma^2=\operatorname{Var}(C'\hat{\beta}).\end{aligned}$$

这就证明了 $C'\hat{\beta}$ 的线性最优性. 又假定 $L_1 \neq L_2$, $L_1'Y$、$L_2'Y$ 为 $C'\beta$ 的两个无偏估计, 令 $L=\frac{1}{2}(L_1+L_2)$, 则 $L'Y$ 也是 $C'\beta$ 的无偏估计. 由严凸函数的性质知, $L'L<\frac{1}{2}(L_1'L_1+L_2'L_2)$, 从而 $\mathrm{Var}L'Y<\frac{1}{2}(\mathrm{Var}L_1'Y+\mathrm{Var}L_2'Y)$, 这说明 $L_1'Y$ 与 $L_2'Y$ 不可能同时为 BLUE, 从而证明了 BLUE 的唯一性.

由(6.14), 我们有

$$\begin{aligned}\mathrm{cov}(C'\hat{\beta},\ D'\hat{\beta}) &= \mathrm{cov}(C'(X'X)^-X'Y,\ D'(X'X)^-X'Y)\\ &= C'(X'X)^-X'X(X'X)^-D\sigma^2\\ &= C'(X'X)^-D\sigma^2.\end{aligned}$$

这就证明了(6.12)式; (6.13)式是(6.12)的直接结果. 定理证毕.

在本段的最后, 我们指出: 下式为 σ^2 的无偏估计:

$$\begin{aligned}S^2 &= \frac{1}{n-r}[Y-X\hat{\beta}]'[Y-X\hat{\beta}]\\ &= \frac{1}{n-r}[Y-X(X'X)^-X'Y]'[Y-X(X'X)^-X'Y]\\ &= \frac{1}{n-r}Y'[I-X'(X'X)^-X]Y,\end{aligned} \tag{6.15}$$

其中 $r=R(X)$. 注意, 由广义逆的性质(vii)(b), 上式与广义逆 $(X'X)^-$ 的选择无关.

证 由广义逆的性质(vii)、(iii), 并注意 $\mathrm{Var}(Y)=\sigma^2 I_n$, 于是有

$$\begin{aligned}ES^2 &= \frac{1}{n-r}E[Y'(I-X(X'X)^-X')Y]\\ &= \frac{1}{n-r}E[(Y-X\beta)'(I-X(X'X)^-X')(Y-X\beta)]\\ &= \frac{1}{n-r}\mathrm{trace}(I-X(X'X)^-X')\sigma^2\\ &= \frac{1}{n-r}[n-\mathrm{trace}(X(X'X)^-X')]\sigma^2\\ &= \frac{1}{n-r}[n-R(X)]\sigma^2=\sigma^2.\end{aligned}$$

三、正态线性模型

假定在线性模型(6.5)中，$e=(e_1, \cdots, e_n)'$服从正态分布$N(0 1_n, I_n\sigma^2)$，$0<\sigma<\infty$. 则称其为正态线性模型. 此时，条件(6.6)仍然成立，故正态线性模型是 Gauss-Markov 模型的特殊情况，从而上面得出的关于 Gauss-Markov 的一切结论在这里仍然成立. 不过，在正态假设下，还可以得出更好的结果来.

在正态假设下，$Y=(y_1, \cdots, y_n)'$的分布密度为

$$
\begin{aligned}
& f(y_1, \cdots, y_n; \beta, \sigma^2) \\
& =\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left\{-\frac{1}{2\sigma^2}(Y-X\beta)'(Y-X\beta)\right\} \\
& =Q(\theta)\exp\left(\sum_{i=0}^{p}\theta_i T_i(Y)\right),
\end{aligned}
\tag{6.16}
$$

其中$\theta_0=-\dfrac{1}{2\sigma^2}$，$\theta_i=\beta_i/\sigma^2 \ (i=1, \cdots, p)$，

$$
Q(\theta)=\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left(-\frac{1}{2\sigma^2}(X\beta)'(X\beta)\right),
$$

$T_i=X_i'Y$，此处X_i表示X的第i列$(i=1, \cdots, p)$. 可见，(6.16)为一指数族，自然参数空间$\Theta=\{(\theta_0, \theta_1, \cdots, \theta_p): \theta_0<0, -\infty<\theta_i<\infty, i=1, \cdots, p\}$，因有内点，根据定理 3.1，$T=(T_0, \cdots, T_p)$为完全充分统计量. 假定$C'\beta$可估，那么其$LS$估计

$$
C'\hat{\beta}=C'(X'X)^{-}X'Y=C'(X'X)^{-}(T_1, \cdots, T_p)'
$$

是T的函数，从而由广义 Blackwell-Rao-Lehmann-Scheffe 定理，有以下结果：

定理6.4 在正态线性模型中，可估函数$C'\beta$的LS估计$C'\hat{\beta}$为其唯一的 UMVUE.

注1 σ^2的无偏估计$\dfrac{1}{n-r}(Y-X\hat{\beta})'(Y-X\hat{\beta})$不能表为$T$的函数，故它不是$\sigma^2$的 UMVUE.

注2 在一般的 Gauss-Markov 模型里，BLUE 未必为 UMVUE. 请看下例：

例 6.1 设 $y_i=\theta+e_i$ $(i=1, \cdots, n)$. 此处 $e_1, \cdots, e_n$ 为 *iid.* 随机变量, $e_1\sim R(-\theta, \theta)$, $0<\theta<\infty$. 这是一个 Gauss-Markov 模型. 此处设计矩阵 $X=\mathbb{1}_n$, $\beta=\theta$. 根据定理 6.2 的 (iii), θ 可估, 且 θ 的 LS 估计为

$$\hat{\theta}=(X'X)^{-1}X'Y=(\mathbb{1}_n'\mathbb{1}_n)^{-1}\mathbb{1}_n'Y=\frac{1}{n}\sum_1^n y_i=\overline{Y}.$$

但因 $Y_1, \cdots, Y_n$ 为 *iid.* 随机变量, $Y_1\sim R(0, 2\theta)$, $0<\theta<\infty$, 由前面例 4.1 知, θ 的 UMVUE 为 $\hat{\theta}_*=\frac{1}{2}\left(1+\frac{1}{n}\right)y_{(n)}$, 此处 $y_{(n)}=\max(y_1, \cdots, y_n)$. 可见, 当 $n>1$ 时, $\overline{Y}$ 不是 θ 的 UMVUE. 事实上, 利用(6.12)式, 可算得

$$\operatorname{Var}(\hat{\theta})=\frac{\theta^2}{3n}.$$

另一方面, 又容易算出 $\operatorname{Var}(\hat{\theta}_*)=\frac{\theta^2}{n(n+2)}$, 可见当 $n>1$ 时, $\operatorname{Var}_\theta(\hat{\theta}_*)<\operatorname{Var}_\theta(\hat{\theta})$.

不过, 由于 BLUE, 即 LS 估计求解方便, 在实际中被广泛采用.

四、广义 Gauss-Markov 模型

若把关于 Gauss-Markov 模型的条件(6.6)修改为

$$\begin{cases} E(e)=0, \\ \operatorname{Var}(e)=\sigma^2 G. \end{cases} \tag{6.6$'$}$$

此处 G 为已知的非负定 n 阶矩阵. 满足(6.6)$'$ 的模型(6.5)称为广义 Gauss-Markov 模型. 当 G 为正定时, 容易把广义 Gauss-Markov 模型变为 Gauss-Markov 模型来研究. 事实上, 作线性变换

$$\widetilde{Y}=G^{-\frac{1}{2}}Y, \quad \widetilde{Y}=(\tilde{y}_1, \cdots, \tilde{y}_n)'.$$

我们便得到新的线性模型

$$\tilde{y}_i=\tilde{x}_{i1}\beta_1+\cdots+\tilde{x}_{ip}\beta_p+\tilde{e}_i \quad (i=1, \cdots, n) \tag{6.5$'$}$$

或记为

$$\widetilde{Y}=\widetilde{X}\beta+\tilde{e},$$

其中 $\widetilde{X}=(\tilde{x}_{ij})=G^{-\frac{1}{2}}X$, $\tilde{e}=(\tilde{e}_1, \cdots, \tilde{e}_n)'=G^{-\frac{1}{2}}e$. 根据(6.6)′, 易推得:

$$\begin{cases} E(\tilde{e})=0, \\ \mathrm{Var}(\tilde{e})=\sigma^2 I_n. \end{cases}$$

即(6.5)′为 Gauss-Markov 模型. 把定理6.2、6.3、6.4应用于模型(6.5)′, 然后再转换回到原来的模型, 便可得到如下结果:

当广义 Gauss-Markov 模型里的误差向量 e 的协方差阵 G 为正定时, 则有:

(i) $C'\beta$ 可估的充要条件是 $C\in\mu(X')$;

(ii) 若 $C'\beta$ 可估, 则 $C'\tilde{\beta}$ 为其 BLUE, 其中

$$\tilde{\beta}=(X'G^{-1}X)^{-}X'G^{-\frac{1}{2}}Y;$$

(iii) 若 $C'\beta$、$D'\beta$ 可估, 则有

$$\mathrm{cov}(C'\tilde{\beta},\ D'\tilde{\beta})=C'(X'G^{-1}X)^{-}D\sigma^2;$$

(iv) $\dfrac{1}{n-r}[(Y-X\tilde{\beta})'G^{-1}(Y-X\tilde{\beta})]$ 为 σ^2 的无偏估计, 其中 $r=R(X)$;

(v) 当 e 为正态时, $C'\tilde{\beta}$ 为可估函数 $C'\beta$ 的唯一的 UMVUE.

但是, 当 G 为奇异时, 情况就复杂了, 上述结果都要作适当的修正, 关于这方面的论述可参看[10].

问题与习题

1. 设 $X\sim B(n, p)$, $0<p<1$. 实函数 $g(p)$ 可估的充要条件是: $g(p)$ 为 p 的 $k(\leqslant n)$ 次多项式.

2. 设 $X_1, \cdots, X_n$ 为 $iid.$, $X_1\sim\mathscr{P}(\lambda)$, $0<\lambda<\infty$, 记 $T=\sum\limits_1^n X_i$. 则 $g(\lambda)$ 存在依赖于 T 的无偏估计的充要条件是: $g(\lambda)$ 为一收敛级数

$$\sum_{k=0}^{\infty}a_k\lambda^k,\ 0<\lambda<\infty.$$

3. 设 $X_1, \cdots, X_n$ 为抽自 $N(a, \sigma^2)$ 的 iid 样本, $-\infty<a<\infty$, $0<\sigma<\infty$. 证明 $g(a, \sigma)=|a|$ 不可估.

4. 设$g(\theta)$为在某分布族$(P_\theta,\ \theta\in\Theta)$下的可估实函数，$\hat{g}^*(x)$为其UMVUE. 则对$g(\theta)$的任一无偏估计$\hat{g}(x)$，当$\mathrm{Var}_\theta\hat{g}<\infty$，$\theta\in\Theta$时，$\hat{g}^*$与$\hat{g}$的相关系数非负. $\hat{g}$也为$g(\theta)$的UMVUE的充要条件是: $\hat{g}$与$\hat{g}^*$的相关系数为1.

5. 设$(P_\theta,\ \theta\in\Theta)$为样本空间$(\mathscr{X},\ \mathscr{B}_x)$上的分布族. $\mathscr{D}_0=\{S(X)\}$某一随机变量族，满足条件: 若S_1、$S_2\in\mathscr{D}_0\Rightarrow a_1S_1+a_2S_2\in\mathscr{D}_0$，其中$a_1$、$a_2$是任二实数. $\hat{g}_0(X)$为一方差有限的随机变量. 记$\mathscr{D}=\{\hat{g}=\hat{g}_0+S: S\in\mathscr{D}_0\}$. 试证: $\hat{g}_0$在$\mathscr{D}$中方差一致最小$\Leftrightarrow$对任一$S\in\mathscr{D}_0$，若$\mathrm{Var}_\theta S<\infty$，则有$\mathrm{cov}_\theta(\hat{g}_0,\ S)=0$.

6. 设$X_1,\ \cdots,\ X_n$为$iid.$随机变量，$X_1\sim N(a,\ \sigma^2)$，$-\infty<a<\infty$，$0<\sigma<\infty$. 求a/σ^2、$e^{-\frac{1}{2}\sigma^2}$的UMVUE.

7. 证明: 例2.7中的$(X_{(1)},\ T)$为完全统计量.

8. 证明: β分布族

$$f(x;\theta_1,\ \theta_2)=\frac{\Gamma(\theta_1+\theta_2)}{\Gamma(\theta_1)\Gamma(\theta_2)}x^{\theta_1-1}(1-x)^{\theta_2-1}$$

$$(0<x<1,\ 0<\theta_1,\ \theta_2<\infty)$$

为完全分布族.

9. 证明: 负二项分布

$f(x;p)=\binom{r+x-1}{x}p^r(1-p)^x\ (x=0,\ 1,\ \cdots,\ r>0$已知，$0<p<1)$是完全分布族. 又设$X_1,\ \cdots,\ X_n$为取自此分布族的$iid.$样本. 证明$T=\sum_1^n X_i$为完全充分统计量，并求$g(p)=p$的UMVUE.

10. 设有两个互相独立的简单样本: $X_1,\ \cdots,\ X_n$，$X_1\sim N(a,\ \sigma^2)$和$Y_1,\ \cdots,\ Y_n$，$Y_1\sim N(a,\ 2\sigma^2)\ (-\infty<a<\infty,\ 0<\sigma<\infty)$. 试利用联合样本$(X_1,\ \cdots,\ X_n,\ Y_1,\ \cdots,\ Y_n)$来求$a$的UMVUE.

11. 设$X\sim N(a,\ \sigma^2)$，证明$E[\Phi(X)]=\Phi\left(\frac{a}{\sqrt{1+\sigma^2}}\right)$. 且利用此事实证明: 若$X_1,\ \cdots,\ X_n$为抽自$N(\theta,\ 1)\ (-\infty<\theta<\infty)$的简单样本，$\tau$为给定实数，则$\Phi(\tau-\theta)$的UMVUE为$\Phi\left(\frac{\tau-\overline{X}}{\sqrt{1-\frac{1}{n}}}\right)$.

12. 设$-\infty\leqslant a<b\leqslant\infty$，$h(x)>0$为定义于$(a,\ b)$上的可测函数，

$$0<\int_c^d h(x)dx<\infty,$$

对任何$[c,\ d]\subset(a,\ b)$，称分布族

$$f_{\theta_1,\theta_2}(x)=\begin{cases}K(\theta_1,\ \theta_2)h(x), & 当\ x\in(\theta_1,\ \theta_2)\\ 0, & 其它\end{cases}\quad (a<\theta_1<\theta_2<b)$$

为双边截断型分布族，这里 $K(\theta_1,\ \theta_2)=\left[\int_{\theta_1}^{\theta_2}h(x)dx\right]^{-1}$. 设 $X_1,\cdots,X_n(n\geqslant 2)$ 为抽自此分布族的简单样本. (i)证明 $(X_{(1)}X_{(n)})$ 为完全充分统计量; (ii)求 $P\{(X_1,\ X_2)\in A|x_{(1)},\ x_{(n)}\}$，其中 A 为任一二维 Borel 集; (iii)设 $g(\theta_1,\ \theta_2)$ 为一待估函数，给出它具有依赖于 X_1、X_2 的无偏估计的充要条件; (iv)当(iii)中的充分条件满足时，求 $g(\theta_1,\ \theta_2)$ 的 UMVUE.

13. 设 $T:(\mathscr{X},\ \mathscr{B}_x,\ P_\theta,\ \theta\in\Theta)\to(\mathscr{T},\ \mathscr{B}_T,\ P_\theta^T,\ \theta\in\Theta)$ 为一统计量. 假若对任一有界 $\mathscr{B}_T$ 可测函数 $\phi(t)$，由"$E_\theta\phi(t)=0$, 对每一 $\theta\in\Theta$"可推出"$\phi(t)=0$, a. s. P_θ^T，对每一 $\theta\in\Theta$"，则称 T 为有界完全的.

13.1. 若分布族为

$$P_\theta[X=x]=\begin{cases}\theta, & 当\ x=-1\\ (1-\theta)\theta^x, & 当\ x=0,\ 1,\ 2,\ \cdots\end{cases}\quad (0\leqslant\theta\leqslant 1)$$

证明 $T(X)=X$ 是有界完全的，但不是完全的. 并证此分布族不存在完全充分统计量.

13.2. 若 $(X_1,\ \cdots,\ X_n)\sim N(\theta\mathbb{1}_n,\ \theta^2I_n)$, $\theta>0$. 则 $T(X)=(\bar{X},\ S^2)$ 非有界完全. 其中 $\mathbb{1}_n$ 表示分量全为 1 的 n 维向量.

13.3. 设 $X_1,\ \cdots,\ X_n$ 为 $iid.$ 随机变量. $X_1\sim R\left(\theta-\frac{1}{2},\ \theta+\frac{1}{2}\right)$, $-\infty<\theta<\infty$, $T(X)=(X_{(1)},\ X_{(n)})$ 非有界完全.

若把分布改为 $R(\theta,\ 2\theta)$, $\theta>0$, 有同样结论.

14. 设 $T:(\mathscr{X},\ \mathscr{B}_x,\ P_\theta,\ \theta\in\Theta)\to(\mathscr{T},\ \mathscr{B}_T,\ P_\theta^T,\ \theta\in\Theta)$ 为一统计量. 若对某一 $\alpha\in(0,\ 1)$，有

$$\text{“}\int_{\mathscr{X}}\phi(x)dP_\theta(x)=\alpha,\ 0\leqslant\phi(x)\leqslant 1,\ \theta\in\Theta\text{”}$$

$$\Rightarrow\text{“}E_\theta(\phi(X)\,|\,t)=\alpha,\ \text{a.s.}P_\theta^T,\ \theta\in\Theta\text{”}.$$

则 T 是有界完全的.

15. 设 T 为如上题定义的统计量有界完全. 若 $V(X)$ 是一分布与参数 θ 无关的随机变量，则 T 与 V 独立(在每个分布 P_θ 下, $\theta\in\Theta$).

16. 设 $X_1,\ \cdots,\ X_n$ 为 $iid.$ 随机变量, $X_1\sim N(0,\ \sigma^2)$, $0<\sigma<\infty$. 证明 $\sum_1^n X_i^2$ 与 $\sum_1^n\lambda_iX_i^2/\sum_1^n X_i^2$ 独立，这里 $\lambda_1,\ \cdots,\ \lambda_n$ 为已知实数.

17. 设 $X_1,\ \cdots,\ X_n$ 为 $iid.$ 随机变量, $X_1\sim N(a,\ \sigma_1^2)$, $-\infty<a<\infty$, $0<\sigma_1<\infty$; $Y_1,\ \cdots,\ Y_n$ 也为 $iid.$ 随机变量, $Y_1\sim N(b,\ \sigma_2^2)$, $-\infty<b<\infty$, $0<\sigma_2<\infty$. 此二样本互相独立. 记

$$\overline{X}=\frac{1}{n}\sum_{1}^{n}X_i, \qquad \overline{Y}=\frac{1}{n}\sum_{1}^{n}Y_i,$$

$$S_X^2=\frac{1}{n-1}\sum_{1}^{n}(X_i-\overline{X})^2, \quad S_Y^2=\frac{1}{n-1}\sum_{1}^{n}(Y_i-\overline{Y})^2,$$

$$r=\frac{\sum_{1}^{n}(X_i-\overline{X})(Y_i-\overline{Y})}{\sqrt{\sum_{1}^{n}(X_i-\overline{X})^2\sum_{1}^{n}(Y_i-\overline{Y})^2}}.$$

证明：r 与$(\overline{X}, \overline{Y}, S_X^2, S_Y^2)$独立.

18. 设有分布族

$$f_\theta(x)=\begin{cases} e^{-(x-\theta)}, & \text{当 } x>\theta \\ 0, & \text{当 } x\leqslant\theta \end{cases} \quad (0<\theta<\infty)$$

$X_1, \cdots, X_n$ 为 $iid.$ 样本. 求 $\exp\sqrt{\theta}$、$\log\theta$ 的 UMVUE.

19. 设 $X_1, \cdots, X_n$ 为 $iid.$ 随机变量, $X_1\sim N(\theta, 1)$, $-\infty<\theta<\infty$. 求 $P_\theta[\max(X_1, \cdots, X_k)<\tau]=[\Phi(\tau-\theta)]^k$ 的 UMVUE.

20. 设 $X_1, \cdots, X_n$ 为 $iid.$ 样本, $X_1\sim R(\theta_1, \theta_2)$, $-\infty<\theta_1<\theta_2<\infty$, 求 θ_1, θ_2 的 UMVUE.

21. 设$(X_1, Y_1), \cdots, (X_n, Y_n)$为 $iid.$ 随机向量,

$$(X_1, Y_1)\sim N\left(\begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix}, \begin{bmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{bmatrix}\right),$$

$$-\infty<\mu_1, \mu_2<\infty, \ 0<\sigma_1, \sigma_2<\infty, \ -1\leqslant\rho\leqslant1.$$

求相关系数 ρ 的 UMVUE.

22. 设 $X_1, \cdots, X_n$ 为抽自 I 型截断族的简单样本. 求分布函数

$$g(z;\theta)=\begin{cases} 0, & \text{当 } z\leqslant\theta \\ K_1(\theta)\int_\theta^z h_1(x)dx, & \text{当 } \theta<z<b \\ 1, & \text{当 } b\leqslant z \end{cases}$$

的 UMVUE. 其中 z 为已知数.

23. 设 $X_1, \cdots, X_n$ 为抽自 I 型截断族的简单样本. 并设 $b=\infty$. $g(\theta)$ 是一待估函数. 对任一$[c, d]\subset(a, \infty)$, $g(\theta)$在其上绝对连续. 对于充分大的 d, $g(\theta)$在(a, d)上的 UMVUE 为 $\hat{g}_d(X_{(1)})$, 并满足条件 $|\hat{g}_d(x)|\leqslant G(x)$ $(a<x<\infty)$, 这里 G 满足条件: $E_\theta G(X_{(1)})<\infty$, 对一切的 $\theta\in(a, \infty)$. 证明: $\lim\limits_{d\to\infty}\hat{g}_d(x)$存在, $a<x<\infty$, 记其极限为 $\hat{g}_\infty(x)$, 则 $\hat{g}_\infty(X_{(1)})$为 $g(\theta)$在(a, ∞)上的 UMVUE. 并把此结果推广到 II 型截断族上.

24. 设 $X_1, \cdots, X_n$ 为 $iid.$ 随机变量. X_1 具有转移型参数族密度 $f(x_1-\theta)dx_1$, $-\infty<\theta<\infty$. $U(X_1, \cdots, X_n)$为转移型统计量. 若 $E_0U=\eta^0$

有限，则 $U-\eta_0$ 为 θ 的无偏估计.

以下各题考虑线性模型 $Y=X\beta+e$, $Ee=0$, $\mathrm{Var}(e)=\sigma^2G=\sigma^2(\sigma_{ij})_{n\times n}$, $Y=(y_1, \cdots, y_n)'$, $X=(x_{ij})_{n\times p}$, $\beta=(\beta_1, \cdots, \beta_p)'$, $e=(e_1, \cdots, e_n)'$, 并记 $S=X'X$.

25. 设 $G=I_n$, 又设 $P'\beta$ 可估，若 $L_0\in\mu(X)$, 且使 $P=X'L_0$, 则 $L_0'Y$ 为 $P'\beta$ 的 LBUE.

26. 设 $G=I_n$, $L'Y$ 为 $E(L'Y)$的 LBUE$\Leftrightarrow$当 $D'X=0$ 时，有 $L'D=0$.

27. 设 $G=I_n$, $P'\beta$ 可估，$a'Y$ 为 $P'\beta$ 的线性无偏估计 $\Leftrightarrow E_{(\beta,\sigma)}\{(a'Y-P'\beta)^2/\sigma^2\}$对一切 σ、β 有界.

28. B 为一对称的 n 阶方阵，有 $E(e'Be)=\sigma^2tr(BG)$; 又若 e 为正态的，则 $\mathrm{Var}(e'Be)=2\sigma^4tr((BG)^2)$.

29. A 为 P 阶对称方阵. 证明: $\hat{\beta}'A\hat{\beta}-\hat{\sigma}^2tr(S^{-1}A)$为 $\beta'A\beta$ 的无偏估计，其中 $\hat{\beta}=S^{-1}X'Y$, $\hat{\sigma}^2=(Y-X\hat{\beta})'(Y-X\hat{\beta})$, 并假定 S 满秩.

30. 设 $T_1, \cdots, T_k$为单参数 θ 的无偏估计，并且 $\mathrm{cov}(T_i, T_j)=\sigma_{ij}(i, j=1, \cdots, k)$. 求 $T_1, \cdots, T_k$ 的线性组合组成的 θ 的无偏估计，使其方差达到最小.

又若 $\sigma_{ij}=\sigma_i^2\delta_{ij}$ $(i, j=1, \cdots, k)$. 证明当 $C_i=\sigma_i^{-2}\Big/\sum_1^k\sigma_j^{-2}(i=1, \cdots, k)$ 时，$\sum_1^k C_iT_i$ 为 θ 的最优线性组合无偏估计，并求其方差.

参 考 文 献

[1] Halmos, P. R. (1946): The theory of unbiased estimation A. M. S. (17) 37～43.

[2] Lehmann, E. L., Scheffé, H. (1950): Completeness, Similar regions and unbiased estimation, Part I, Sankhya, 10:305～340.

[3] Kolmoyorov, A. N. (1950): Unbiased estimates, Izvestia Akad,Nauk SSSR, Seriya Mathmaticeskava, 14: 303～326 (Amer. Math. Soc.Translation No. 98).

[4] Epstein, B. Sobel, M. (1954): Some theorems relevant to life testing from an exponential distribution, A. M. S., 25: 273～381.

[5] Lieberman, G. J., Resnikoff, G. J. (1955): Sampling plans for inspection by variables, Jour. Amer. Statist. Assoc. 50: 457～516.

[6] Washio, Y., Morimoto, H. And Ikeda, (1956): Unbiased estimation based on sufficient statistics, Bull of Math. Stat. (6) 69～94.

[7] Fraser, D. A. S. (1957): Nonparametric Methods in Statistics, John Wiley, New York.

[8] Tate, R. F. (1959): Unbiased estimation: Functions of location and scale parameters, Ann. Math. Statist. 30: 341~366.

[9] Zyskind, O. (1967): On Canonical forms nonnagative covariance matrices and best and simple least-squares linear estimators inlinear models. A. M. S. (38) 1092~1109.

[10] Rao, C. R. (1973): Linear statistical Inference and Its Applications. Wiley, New York.

[11] Mood, A. M., Graybill, F. A., and Boes, D. C. (1974): Introduction to the Theory of Statistics. McGraw-Hill Book co. New York (p. 330).

[12] Jagdish K. Patel, Kapadia, C. H. and Owen, D. B. (1976): Handbook of Statistical Distributions.

[13] Zacks, S. (1971): The Theory of Statistical Inference. Wiley, New York.

[14] 卢昆亮、赵林城(1980):《双边截断参数的无偏方差最小估计》载于《科学通报》, 3, 102~104. 全文载于《应用数学学报》, 1982, 2:177~185.

[15] 陈桂景、陈希孺(1981):《UMVUE 不存在的一种分布族》载于《科学通报》. 15: 902~904. 全文载于《数学研究与评论》, 1984, 3:93~98.

第三章　C-R型不等式

前面我们已讨论了参数的一致最小方差无偏估计(UMVUE). 本章主要讨论分布族满足各种正则条件下的无偏估计方差的下界以及达到这些下界的充要条件. 最早是 Cramér 和 Rao 分别在 1945 和 1946 年发现了一个下界[1]、[2]，现称为 C-R 不等式(或下界)，随后 Bhattacharyya 于 1946 年把它推广为更精确的下界[3]. 这些下界仅适用于有共同支撑(密度函数不为零的样本点区域与参数无关)的情况. 其它情况，本章也将涉及. 这些不等式在参数估计理论中起着很重要的作用，因此，单独用一章来叙述.

本章用$(\mathscr{X}, \mathscr{B}_x)$表示样本空间，$\mathscr{P}=\{P_\theta, \theta\in\Theta\}$ 表示其分布族，令μ是$(\mathscr{X}, \mathscr{B}_x)$上的$\sigma$有限测度. 我们假设 $\mathscr{P}\ll\mu$，记

$$f(x, \theta)=\frac{dP_\theta}{d\mu}.$$

§1　单参数的C-R下界

一、C-R 下界

我们假设分布族$\mathscr{P}$满足以下正则条件:

i) Θ 是实轴 R 中的开区间(有穷或无穷);

ii) $f(x, \theta)$有共同支撑，对一切 $x\in\mathscr{X}$, $\theta\in\Theta$, $\frac{\partial}{\partial\theta}f(x, \theta)$存在;

iii) 当 $\theta\in\Theta$ 时, $\int\frac{\partial}{\partial\theta}f(x, \theta)d\mu=0$;

iv) 当 $\theta\in\Theta$ 时, $I(\theta)\triangleq E_\theta\left[\frac{\partial}{\partial\theta}\log f(x, \theta)\right]^2>0$;

v）对每个固定的 $x\in\mathscr{X}$, $\frac{\partial}{\partial\theta}f(x,\theta)$ 是 Θ 上的连续函数.

定理1.1 如果分布族 $\mathscr{P}$ 满足上述条件 i)～iv), $g(\theta)$ 是 Θ 上可估并可微的函数, $T(x)$ 满足

vi) $E_\theta[T(x)]=g(\theta)$, 且

$$\int T(x)f(x,\theta)d\mu \text{ 可在积分号下求导.}$$

那么, 下列不等式(C-R 不等式)成立:

$$\mathrm{Var}_\theta\{T\}\geqslant\frac{[g'(\theta)]^2}{E_\theta\left[\frac{\partial}{\partial\theta}\log f(x,\theta)\right]^2}\quad(\text{当 }\theta\in\Theta).\tag{1.1}$$

若分布族还满足 v), $g(\theta)$不恒为常数, 且 $\mathrm{Var}_\theta\{T\}<\infty$. 则(1.1)中的等号对所有的 $\theta\in\Theta$ 成立的充要条件是: $\mathscr{P}$ 为指数型分布族, 即存在 $K\in\mathscr{B}_x$, $\mu(K)=0$, 使

当 $x\bar{\in}K$, $\theta\in\Theta$ 时,

$$f(x,\theta)=c(\theta)h(x)e^{\psi(\theta)T(x)}.\tag{1.2}$$

此处 $c(\theta)>0$, $\psi(\theta)$在 Θ 上可微, $\psi(\theta)$严格单调, $h(x)>0$.

证 记 $Z(x,\theta)=\frac{\partial}{\partial\theta}\log f(x,\theta)$, 当 $\mathrm{Var}_\theta\{T(X)\}=\infty$ 或 $\mathrm{Var}_\theta\{Z(X,\theta)\}=\infty$ 时, (1.1)自然成立. 下设 $\mathrm{Var}_\theta\{T(X)\}<\infty$, $\mathrm{Var}_\theta\{Z(X,\theta)\}<\infty$. 根据 Schwarz 不等式, 对任意随机变量 Y, 若 $0<\mathrm{Var}(Y)<\infty$, 就有

$$\mathrm{Var}_\theta\{T\}\mathrm{Var}_\theta\{Y\}\geqslant[\mathrm{cov}\{T,Y\}]^2.\tag{1.3}$$

因为 $$E_\theta\{Z(X,\theta)\}=\int\frac{\partial}{\partial\theta}f(x,\theta)d\mu=0,$$

$$\mathrm{Var}_\theta\{Z(X,\theta)\}=E_\theta\left\{\frac{\partial}{\partial\theta}\log f(X,\theta)\right\}^2,\tag{1.4}$$

根据条件 vi), 有

$$\begin{aligned}\mathrm{cov}_\theta\{T(X),Z(X,\theta)\}&=E_\theta\{T(X)Z(X,\theta)\}\\&=\int T(x)\frac{\partial}{\partial\theta}f(x,\theta)d\mu\\&=\frac{\partial}{\partial\theta}\int T(x)f(x,\theta)d\mu\end{aligned}$$

$$=g'(\theta). \tag{1.5}$$

取 $Y=Z(X,\ \theta)$，将(1.4)、(1.5)代入(1.3)，即得(1.1)式.

下面证定理的第二部分：由 Schwarz 不等式的等号成立的充要条件可知，(1.1)式中等号成立的充要条件为：存在 $\alpha(\theta)$、$\beta(\theta)$ 不全为零，使

$$\alpha(\theta)[T(x)-g(\theta)]+\beta(\theta)Z(x,\ \theta)=0,\ a.s.P_\theta. \tag{1.6}$$

根据条件 ii)，$a.s.P_\theta$ 与 $a.s.\mu$ 等价. 而 $g(\theta)$ 在 Θ 上不恒为常数，故如果

$$\alpha(\theta)\neq 0,\ \beta(\theta)=0,\ \Rightarrow T(x)=g(\theta),\ a.s.\mu.$$

但 $T(x)$ 与 θ 无关. 这不可能.

如果 $\alpha(\theta)=0,\ \beta(\theta)\neq 0 \Rightarrow Z(x,\ \theta)=0, a.s.\mu$. 这与假设 iv) 矛盾. 这也不可能.

由此推出 $\alpha(\theta)\neq 0,\ \beta(\theta)\neq 0$，从而(1.6)式可表为

$$Z(x,\ \theta)=a(\theta)[T(x)-g(\theta)],\ a.s.\mu. \tag{1.7}$$

或者说：(1.1)式中等号成立的充要条件为：存在 $a(\theta)\neq 0$ 及 $N_\theta\in\mathscr{B}_x$，$\mu(N_\theta)=0$，使当 $x\overline{\in}N_\theta$、$\theta\in\Theta$ 时，(1.7)式成立.

现在利用上述结论来证明定理的第二部分：

充分性：设(1.2)式成立，因为 $c(\theta)>0$、$h(x)>0$ 对(1.2)式两边取对数，即得

$$\log f(x,\ \theta)=\log h(x)+\log c(\theta)+\psi(\theta)T(x)$$
$$(当\ x\overline{\in}K,\ \theta\in\Theta). \tag{1.8}$$

由条件 ii) 及有关 $c(\theta)$、$\psi(\theta)$ 的假设，对(1.8)式两边取微分，得

$$Z(x,\ \theta)=\frac{\partial}{\partial\theta}\log c(\theta)+\psi'(\theta)T(x)$$
$$(当\ x\overline{\in}K,\ \theta\in\Theta). \tag{1.9}$$

因为 $\psi'(\theta)\neq 0$，故(1.7)式成立，充分性得证.

必要性：设(1.1)式等号成立，那么(1.2)式成立. 根据(1.1)式等号成立推得存在 $a(\theta)\neq 0$，μ 零测集 N_θ 当 $x\overline{\in}N_\theta$、$\theta\in\Theta$ 时，

$$Z(x,\ \theta)=a(\theta)\{T(x)-g(\theta)\}. \tag{1.10}$$

由 $f(x,\ \theta)$ 为 $\mathscr{B}_x\times\mathscr{B}_\Theta$ 可测推得 $Z(x,\ \theta)$ 关于 $\mathscr{B}_x\times\mathscr{B}_\Theta$ 可测.

此处 $\mathscr{B}_\Theta$ 为 Θ 上的所有 Borel 集组成的 σ 域，令

$$N=\{(x,\ \theta):(x,\ \theta)\in\mathfrak{x}\times\Theta,$$
$$Z(x,\ \theta)\neq a(\theta)\{T(x)-g(\theta)\}, \tag{1.11}$$

N_θ 是 N 在 θ 处截口. 由于 $g(\theta)$ 不恒为常数且可微，故可取 $\theta_1\neq\theta_2$，使 $g(\theta_1)\neq g(\theta_2)\neq g(\theta)$. 对(1.10)分别在两边取数学期望 $E_{\theta_1}\{\cdot\}$、$E_{\theta_2}\{\cdot\}$，即得

$$a(\theta)=\frac{E_{\theta_1}\{Z(X,\ \theta)\}-E_{\theta_2}\{Z(X,\ \theta)\}}{g(\theta_1)-g(\theta_2)}. \tag{1.12}$$

由于 $Z(x,\ \theta)$ 为 $\mathscr{B}_x\times\mathscr{B}_\Theta$ 可测，故 $E_{\theta_i}\{Z(X,\ \theta)\}\ (i=1,\ 2)$ 为 $\mathscr{B}_\Theta$ 可测，从而 $a(\theta)$ 为 $\mathscr{B}_\Theta$ 可测. 由此推出 N 为 $\mathscr{B}_x\times\mathscr{B}_\Theta$ 可测. 根据 Fibini 定理，得

$$0=\int\mu(N_\theta)dL=\int d\mu\int I_N(x,\ \theta)dL=\int L(N^x)d\mu. \tag{1.13}$$

此处 N^x 表 N 在 x 处的截口，L 表示 Lebesgue 测度. (1.13)式表明存在 $K\in\mathscr{B}_x$，$\mu(K)=0$，使当 $x\bar{\in}K$ 时，$L(N^x)=0$，因此(1.10)可改写为

$$Z(x,\ \theta)=a(\theta)(T(x)-g(\theta))\quad(\text{当 } x\bar{\in}K,\ \theta\bar{\in}N^x). \tag{1.14}$$

由于 $g(\theta)$ 不恒为常数，可取 x_1、$x_2\bar{\in}K$，$T(x_1)\neq T(x_2)$，故

$$Z(x_1,\ \theta)-Z(x_2,\ \theta)=a(\theta)(T(x_1)-T(x_2))$$
$$(\text{当 } \theta\bar{\in}N^{x_1}\cup N^{x_2}). \tag{1.15}$$

由假设 v 推得 $a(\theta)$ 在任何有限区间上可积，再由(1.14)及 g 可微推得 $a(\theta)g(\theta)$ 在任何有限区间上可积. 取定 $\theta_0\in\Theta$，记

$$f(x,\ \theta_0)=h(x),$$

$$\psi(\theta)=\int_{\theta_0}^{\theta}a(t)dt,\quad \log c(\theta)=-\int_{\theta_0}^{\theta}a(t)g(t)dt.$$

由于 $Z(x,\ \theta)$ 关于 θ 连续，根据(1.15)、(1.14)知 $\psi(\theta)$、$c(\theta)$ 可微. 因 $a(t)\neq0$，故 $\psi(\theta)$ 严格单调，$c(\theta)>0$ 是显然的. 由(1.14)有

$$\log f(x,\ \theta)=\log h(x)+\psi(\theta)T(x)+\log c(\theta)\quad(\text{当 } x\bar{\in}K). \tag{1.16}$$

由(1.16)立即推得(1.2).

至此，定理全部证完.

请读者注意：有关 $c(\theta)$、$\psi(\theta)$、$h(x)$ 的性质不必作为充分条件，它可由假设条件推出．定理 1.1 表明：在正则条件下，除指数族分布外，都不能处处达到 C-R 下界．即使指数族分布，也只有很特殊一类 $g(\theta)$ 才有处处达到下界的无偏估计．

推论 1.1 在满足定理条件的指数型分布族中，当且仅当

$$g(\theta)=-\alpha[\psi'(\theta)]^{-1}\frac{c'(\theta)}{c(\theta)}+\beta=\alpha E_\theta\{T(X)\}+\beta \qquad (1.17)$$

时，才有处处达到 C-R 下界的无偏估计 $\hat{g}(x)=\alpha T(x)+\beta$.

证 由指数族的性质可知

$$E_\theta\{T(X)\}=-[\psi'(\theta)]^{-1}\frac{c'(\theta)}{c(\theta)};$$

$$\mathrm{Var}_\theta\{T(X)\}=(\psi'(\theta))^{-2}E_\theta\{Z^2(X,\theta)\}=(\psi'(\theta))^{-2}I(\theta).$$

而

$$\begin{aligned}g'(\theta)&=\alpha\frac{\partial}{\partial\theta}E_\theta\{T(X)\}\\&=\alpha E_\theta\left\{T(X)\left[\psi'(\theta)T(X)+\frac{c'(\theta)}{c(\theta)}\right]\right\}\\&=\alpha(\psi'(\theta))^{-1}E_\theta\{Z^2(X,\theta)\}=\alpha(\psi'(\theta))^{-1}I(\theta).\end{aligned}$$

故 $$\{g'(\theta)\}^2I^{-1}(\theta)=\alpha^2(\psi'(\theta))^{-2}I(\theta)=\mathrm{Var}(\hat{g}(X)).$$

由此，充分性得证．

必要性：由定理 1.1 的证明过程可知道，存在 $a(\theta)\neq 0$，$K\in\mathscr{B}_x$，$\mu(K)=0$，使

$$Z(x,\theta)=a(\theta)[\hat{g}(x)-g(\theta)] \quad (\text{当 } x\overline{\in}K,\ \theta\in\Theta).$$

而此时 $$Z(x,\theta)=\frac{c'(\theta)}{c(\theta)}+\psi'(\theta)T(x),$$

从而

$$\hat{g}(x)=\frac{c'(\theta)}{a(\theta)c(\theta)}+g(\theta)+a^{-1}(\theta)\psi'(\theta)T(x). \qquad (1.18)$$

由于 $T(x)$ 是非退化的，$\hat{g}(x)$ 与 θ 无关，推得

$$\psi'(\theta)a^{-1}(\theta)=\alpha=\text{常数};$$

$$a^{-1}(\theta)c^{-1}(\theta)c'(\theta)+g(\theta)=\beta=\text{常数}. \qquad (1.19)$$

由(1.18)、(1.19)推得

$$\hat{g}(x)=\alpha T(x)+\beta.$$

推论至此证完.

利用(1.1)还可给出一个有偏估计的均方误差的下界.

推论 1.2 假设分布族 $\mathscr{P}$ 满足正则条件 i)～iv), $g(\theta)$ 在 Θ 上可微, $T(x)$ 是 $g(\theta)$ 的估计, 而且满足假设 vi), 那么

$$E_\theta\{T(X)-g(\theta)\}^2 \geqslant b^2(\theta)+(g'(\theta)+b'(\theta))^2 I^{-1}(\theta), \quad (1.20)$$

式中 $b(\theta)=g(\theta)-E_\theta\{T(X)\}$.

证 利用(1.1)式即可推得.

应当指出: 如正则条件 v) 不成立, 则(1.2)式仅是(1.1)式等号成立的充分条件, 而非必要条件.

例 1.1(Joshi)[6] 取 $0<\alpha<1$, β 为

$$\int_\alpha^\beta (t^2-1)e^{-\frac{t^2}{2}}dt=0$$

的根. 记

$$A(x)=I_{[\alpha,\beta]}(|x|)+1.$$

考虑密度函数

$$f(x,\ \theta)=f(x-\theta)=cA(|x-\theta|)e^{-\frac{(x-\theta)^2}{2}} \quad (\text{当 } \theta\in\Theta=R),$$

其中 $c^{-1}=\sqrt{2\pi}+2\int_\alpha^\beta e^{-\frac{t^2}{2}}dt$.

由于 $f(x)$ 关于原点对称, 故 x 是 θ 的无偏估计.

$$Z(x,\ \theta)=\frac{\partial}{\partial\theta}\log f(x,\ \theta)=x-\theta \quad (\text{当 } x\neq\theta\pm\alpha,\ \theta\pm\beta).$$

$$E_\theta(X-\theta)^2=c\int_{-\infty}^{\infty}x^2A(|x|)e^{-\frac{x^2}{2}}dx=1=E_\theta(Z^2(X,\ \theta))$$
$$=I^{-1}(\theta).$$

这说明 C-R 不等式的等号成立. 但 $f(x,\ \theta)$ 不是指数型分布族. 因为 $Z(x,\ \theta)$ 不是 θ 的连续函数.

二、有 效 率

对于正则条件下的分布族, 因为有 C-R 下界, 人们就用它建立一个衡量无偏估计好坏的标准. 提出有效率的概念. 设 $g(\theta)$ 存在无偏估计, $\hat{g}$ 是 $g(\theta)$ 一个无偏估计, 我们称

$$(g'(\theta))^2 I^{-1}(\theta)/\mathrm{Var}_\theta(\hat{g})$$

为估计 $\hat{g}(x)$ 的有效率；如果它等于1，则称 $\hat{g}$ 为 $g(\theta)$ 的有效估计.

这种有效率的定义是值得商榷的. 一些满足正则条件的分布族，可估函数 $g(\theta)$ 的所有无偏估计的方差有它本身的下界，例如 UMVUE. 但它的效率不常为1. 这就不合理了.

例 1.2 设 $x_1, x_2, \cdots, x_n$ 为 n 个 iid. 随机变量，具有 Poisson 分布 $\mathscr{P}(\lambda)$. 我们要估计 $g(\lambda)=e^{-\lambda}$. 众所周知 $\sum_{i=1}^{n} x_i \sim \mathscr{P}(n\lambda)$ 而且是 λ 的完全充分统计量. 现取 $g(\lambda)$ 的无偏估计

$$\hat{g}_n=\left(1-\frac{1}{n}\right)^{\sum_1^n x_i},$$

它的方差为 $e^{-2\lambda}(e^{\frac{\lambda}{n}}-1)$. 由广义 B-R-L-S 定理知，$\hat{g}_n$ 是 $g(\lambda)$ 的 UMVUE. 但 C-R 下界不难算得为 $\frac{\lambda}{n}e^{-2\lambda}$，它小于 $\mathrm{Var}_\lambda(\hat{g}_n)$，这说明 C-R 下界根本达不到.

三、非共同支撑情况

因无偏估计方差的 C-R 下界仅适用于共同支撑的分布族情况，下面介绍几个无偏估计方差的下界：

i) Kiefer 下界[9]. 令 $\Omega_\theta=\{h:\theta+h\in\Theta\}$，$\lambda_1$、$\lambda_2$ 是 Ω_θ 上的任意两个概率测度

$$E_i(h)\doteq\int_{\Omega_\theta} h\,d\lambda_i \quad (i=1,\ 2). \tag{1.21}$$

若 $T(x)$ 为 $g(\theta)$ 的无偏估计，则

$$\mathrm{Var}_\theta(T(X))\geqslant\sup_{\lambda_1,\lambda_2}\frac{[E_1(g(\theta+h))-E_2(g(\theta+h))]^2}{\left\{\int_{f(x,\theta)>0}\left[\int_{\Omega_\theta} f(x,\ \theta+h)\,d(\lambda_1-\lambda_2)\right]^2\times f^{-1}(x,\ \theta)\,d\mu\right\}}. \tag{1.22}$$

利用 Schwatz 不等式即可证明.

若取 λ_2 为 $h=0$ 的退化分布，则(1.22)式成为

$$\mathrm{Var}_\theta(T(X)) \geqslant \sup_{\lambda_1} \frac{[E_1(g(\theta+h))-g(\theta)]^2}{\left\{\int_{f(x,\theta)>0}\left[\int_{\Omega_\theta} f(x,\theta+h)\,d\lambda_1 - f(x,\theta)\right]^2 \times f^{-1}(x,\theta)\,d\mu\right\}}. \tag{1.23}$$

ii) Chapman 和 Robbin 下界[10]. 若在(1.23)式中取 λ_1 为在 h 的退化分布，即得

$$\mathrm{Var}(T(X)) \geqslant \sup_{h\in\Omega_\theta} \frac{(g(\theta+h)-g(\theta))^2}{E_\theta[f(X,\theta th)f^{-1}(X,\theta)-1]^2}. \tag{1.24}$$

令 $h\to 0$，即得到 C-R 不等式.

确有估计可达到(1.23)式的下界(见习题).

§2　Fisher 信息函数

如果分布族 $\mathscr{P}=\{P_\theta,\ \theta\in\Theta\}$ 满足本章§1中的正则条件 i)～iii)，则称

$$I(\theta)=E_\theta\left(\frac{\partial}{\partial\theta}\log f(X,\theta)\right)^2 \quad (\theta\in\Theta)$$

为分布族 $\mathscr{P}$ 的 Fisher 信息函数. 有时更明确地记为 $I^{\mathscr{P}}(\theta)$. C-R 下界仅仅通过 $I(\theta)$ 来表达. 从它的定义可看出，它是描述密度函数随 θ 变化$\left(\text{由 } \frac{\partial}{\partial\theta} \text{ 表达}\right)$的情况. 在某种情况下，表达了分布函数所含参数 θ 的信息量. 首先为 Fisher 所提出，故称为 Fisher 信息函数. 下面我们来探讨一下它的性质和作用:

1° $I(\theta)\geqslant 0$，且与 μ 的选取无关. 也就是说，$I(\theta)$ 确是分布族本身的属性.

设另有 σ 有限测度 μ_1 使得 $\mathscr{P}\ll\mu_1$，由正则条件 ii) 知 $\mathscr{P}\equiv\mu$. 从而 $\mu\ll\mu_1$，记

$$h(x)\triangleq\frac{d\mu}{d\mu_1}.$$

$$f_1(x,\theta)=\frac{dP_\theta}{d\mu_1}=\frac{dP_\theta}{d\mu}\cdot\frac{d\mu}{d\mu_1}=f(x,\theta)h(x),\ a.s.\mu_1;$$

$$\int_{\mathscr{X}_1} f_1(x,\ \theta)\, d\mu_1 = 1.$$

此处 $\mathscr{X}_1 = \{x: h(x) > 0\}$，可视为样本空间，显然

$$\frac{\partial}{\partial\theta}\log f_1(x,\ \theta) = \frac{\partial}{\partial\theta}\log f(x,\ \theta)\quad (\text{当 } \theta\in\Theta,\ x\in\mathscr{X}_1).$$

因此从 $I(\theta)$ 定义得出它与 μ 的选取无关.

2° 若 X、Y 独立，它们的分布族皆满足正则条件 i)～iii)，则 $(x,\ y)$ 的分布族同样满足正则条件 i)～iii)，而且

$$I^x(\theta) + I^y(\theta) = I^{(x,y)}(\theta)\quad (\theta\in\Theta). \tag{2.1}$$

特别是 $X_1,\ X_2,\ \cdots$ 为 iid. 随机序列，那么

$$I^{(x_1, x_2, \cdots x_n)}(\theta) = nI(\theta)\quad (\theta\in\Theta). \tag{2.2}$$

证 根据 1°，不妨设 X、Y 的分布 P_θ^x、P_θ^y 皆有关于 σ 有限测度 μ 的密度 $f(x,\ \theta)$、$g(y,\ \theta)$. 记 $\mu^2 = \mu\times\mu$

$$Z^x(x,\ \theta) = \frac{\partial}{\partial\theta}\log f(x,\ \theta);$$

$$Z^y(y,\ \theta) = \frac{\partial}{\partial\theta}\log g(x,\ \theta).$$

故 $\dfrac{dP_\theta^{x,y}}{d\mu^2} = f(x,\ \theta)g(y,\ \theta)$. 显然满足正则条件 i)～iii). 相应的 $Z^{x,y}(x,\ y;\theta)$ 有

$$Z^{x,y}(x,\ y;\theta) = Z^x(x,\ \theta) + Z^y(y,\ \theta)\quad (\theta\in\Theta).$$

故

$$\begin{aligned} I^{x,y}(\theta) &= E_\theta[Z^x(X,\ \theta) + Z^y(Y,\ \theta)]^2 \\ &= E_\theta(Z^x(X,\ \theta))^2 + E_\theta(Z^y(Y,\ \theta))^2 = I^x(\theta) + I^y(\theta). \end{aligned}$$

利用(2.1)及归纳法即得(2.2)式.

3° 若随机变量 X 及统计量 $T(X)$ 的分布族 $\mathscr{P}$ 及 $\mathscr{P}^T$ 皆满足正则条件 i)～iii) 且满足

iii) 对任意与 θ 无关的 $B\in\mathscr{B}_x$ 皆有

$$\frac{\partial}{\partial\theta}\int_B f(x,\ \theta)\, d\mu = \int_B \frac{\partial}{\partial\theta} f(x,\ \theta)\, d\mu\quad (\theta\in\Theta); \tag{2.3}$$

对任意 $A\in\mathscr{B}_T$(与 θ 无关)皆有

$$\frac{\partial}{\partial\theta}\int_A g(t,\ \theta)\, d\mu^T = \int_A \frac{\partial}{\partial\theta} g(t,\ \theta)\, d\mu^T\quad (\theta\in\Theta). \tag{2.4}$$

此处 $$f(x, \theta)=\frac{dP_\theta^x}{d\mu},\ g(t, \theta)=\frac{dP_\theta^T}{d\mu^T}.$$

在上述条件下就有

$$I^T(\theta)\leqslant I^X(\theta). \tag{2.5}$$

若进一步,正则条件 v) 对 $f(x, \theta)$、$g(t, \theta)$ 成立,则 $I^X(\theta)=I^T(\theta)$ 对一切 $\theta\in\Theta$ 成立的充要条件为: $T(x)$ 是充分统计量.

证 从直观上看,结论是显然的,X 比 $T(x)$ 更多提供 θ 的信息. 如果相等,那只有 $T(x)$ 是充分统计量. 这也是充分统计量的真正含意所在. 现在给以严格的数学证明:

不失一般性,设 $I^X(\theta)<\infty$,记

$$Z^X(x, \theta)=\frac{\partial}{\partial\theta}\log f(x, \theta),\quad Z^T(t, \theta)=\frac{\partial}{\partial\theta}\log g(t, \theta),$$

由假设 iii)′,对任意与 θ 无关的 $B\in\mathscr{B}_T$, 有

$$\begin{aligned}\int_{T^{-1}(B)} Z^T(T(x), \theta)dP_\theta &= \int_B Z^T(t, \theta)dP_\theta^T\\ &=\int_B\frac{\partial}{\partial\theta} g(t, \theta)d\mu^T=\frac{\partial}{\partial\theta}P_\theta^T(B)=\frac{\partial}{\partial\theta}P_\theta^x(T^{-1}(B))\\ &=\frac{\partial}{\partial\theta}\int_{T^{-1}(B)}f(x, \theta)d\mu=\int_{T^{-1}(B)}\frac{\partial}{\partial\theta}f(x, \theta)d\mu\\ &=\int_{T^{-1}(B)}Z^X(x, \theta)dP_\theta.\end{aligned} \tag{2.6}$$

根据(2.6)式即说明

$$Z^T(T(x), \theta)=E(Z^X(x, \theta)\,|\,T(x)),\quad a.s.P_\theta. \tag{2.7}$$

故

$$\begin{aligned}0&\leqslant E_\theta[Z^T(T(x), \theta)-Z^X(x, \theta)]^2\\ &=E_\theta[Z^T(T, \theta)]^2-E_\theta[Z^X(x, \theta)]^2=I^X(\theta)-I^T(\theta).\end{aligned} \tag{2.8}$$

由(2.8)知, 3° 的第一部分已得证,下面证第二部分:

由(2.8)知

$$I^X(\theta)=I^T(\theta)\Leftrightarrow Z^X(x, \theta)=Z^T(T(x), \theta),\quad a.s.P_\theta.$$

必要性: 记

$$N=\{(x, \theta),\ Z^X(x, \theta)\neq Z^T(T(x), \theta)\},$$

象证明定理 1.1 一样,用 Fubini 定理及

$$Z^X(x,\ \theta)=Z^T(T(x),\ \theta),\quad a.s.\ P_\theta$$

知，存在μ零测集K，当$x\bar{\in}K$时，存在Lebesgue零测集N^x，使

$$Z^X(x,\ \theta)=Z^T(T(x),\ \theta)\quad (\text{当}\ x\bar{\in}K,\ \theta\bar{\in}N^x). \qquad (2.9)$$

但根据正则条件v，(2.9)式两边皆关于θ连续．所以N^x为空集．任取$\theta_0\in\Theta$，记$[g(T(x),\ \theta_0)]^{-1}f(x,\ \theta_0)=h(x)$．对(2.9)两边积分后，再取反对数即得

$$f(x,\ \theta)=g(T(x),\ \theta)h(x)\quad (\text{当}\ x\bar{\in}K,\ \theta\in\Theta). \qquad (2.10)$$

因此$T(x)$是θ的充分统计量．

充分性：设T是θ的充分统计量，则对每个$\theta\in\Theta$，存在$N_{1\theta}$，$\mu(N_{1\theta})=0$，及可测函数$\tilde{g}(T(x),\ \theta)$，$h(x)$，使

$$f(x,\ \theta)=\tilde{g}(T(x),\ \theta)h(x)\quad (\text{当}\ x\bar{\in}N_{1\theta}). \qquad (2.11)$$

因此，存在$\mathscr{B}_{\mathscr{T}}$可测函数$\zeta(t)$及$\mu^T$零测集$N_\theta^T$，使得

$$g(t,\ \theta)=\tilde{g}(t,\ \theta)\zeta(t)\quad (\text{当}\ t\bar{\in}N_\theta^T). \qquad (2.12)$$

由正则假设ii)知$g(t,\ \theta)>0$，故$\zeta(t)>0$（当$t\bar{\in}N_\theta^T$）．以(2.12)代入(2.11)，即得

$$f(x,\ \theta)=g(T(x),\ \theta)\zeta^{-1}(T(x))h(x)$$
$$(\text{当}\ x\bar{\in}N_{1\theta}\cup T^{-1}(N_\theta^T)). \qquad (2.13)$$

应用与证明本章定理1.1相似的办法，利用Fubini定理以及$f(x,\theta)$、$g(t,\ \theta)$关于θ连续，则存在μ的零测集（与θ无关）K，当$x\bar{\in}K$，$\theta\in\Theta$时，

$$f(x,\ \theta)=g(T(x),\ \theta)\zeta^{-1}(T(x))h(x). \qquad (2.14)$$

对(2.14)两边关于θ取微分，即得$Z^X(x,\ \theta)=Z^T(T(x),\ \theta)$（当$x\bar{\in}K$）．因此$I^X(\theta)=I^T(\theta)$（当$\theta\in\Theta$）．至此定理证完．

应当指出：若正则条件v不成立时，即使$I^T(\theta)=I^X(\theta)$对一切θ成立，但$T(x)$也不一定是θ的充分统计量．在本章例1.1中，如果独立取样$X_1,\ X_2,\ \cdots,\ X_n$．$I^{(x_1,x_2,\cdots x_n)}(\theta)=nI^{x_1}(\theta)=n$．因例1.1满足C-R不等式成立的条件，故$\bar{x}$的分布族也满足，故有下式成立：

$$I^{\bar{X}}(\theta)\geqslant(E(\bar{X}-\theta)^2)^{-1}=n=I^{x_1,\cdots,x_n}(\theta)\quad (\text{当}\ \theta\in\Theta).$$

由(2.5)式知$I^{\bar{X}}(\theta)\leqslant I^{x_1,\cdots,x_n}(\theta)$，因此$I^{\bar{X}}(\theta)=I^{x_1,\cdots,x_n}(\theta)$，但$\bar{X}$

并不是θ的充分统计量.

§3 单参数的 Bhattacharyya 下界

Bhattacharyya 在 1946 年在更强的正则条件下推广了 C-R 下界,得到了一系列愈来愈大的无偏估计方差的下界.从直观上看,仅用$\log f(x,\theta)$关于θ的偏微分来描述$f(x,\theta)$随θ的变化情况是不够的,如果用各阶偏导数来刻划,显然精确多了.Bhattacharyya 正是从这点出发改进了 C-R 下界,而得到了一串 Bhattacharyya 不等式(简称 Bh 不等式).

一、Bh 不 等 式

正则条件

i*) Θ是R上的开区间(有穷或无穷);

ii*) $f(x,\theta)>0$(当$x\in\mathscr{X}$、$\theta\in\Theta$). $\frac{\partial^i}{\partial\theta^i}f(x,\theta)$存在且有限($i=1,2,\cdots,k;\theta\in\Theta$);

iii*) $\int\frac{\partial^i}{\partial\theta^i}f(x,\theta)d\mu=0\quad(i=1,2,\cdots,k)$;

iv*) $\int\left(\frac{\partial^i}{\partial\theta^i}f(x,\theta)\right)^2f^{-1}(x,\theta)d\mu<\infty\ (\theta\in\Theta;\ i=1,2,\cdots k)$;

v*) $\hat{g}(x)$是$g(\theta)$的无偏估计,$g(\theta)$ k次可微,以及

$$g^{(i)}(\theta)=\int\hat{g}(x)\left[\frac{\partial^i}{\partial\theta^i}f(x,\theta)\right]d\mu=\frac{\partial^i}{\partial\theta^i}\int\hat{g}(x)f(x,\theta)d\mu$$

$$(1\leqslant i\leqslant k;\ \theta\in\Theta;\ g(\theta)\text{不恒为常数}).$$

在介绍定理之前,先叙述一条简单引理:

引理 3.1 设X是m维随机向量,则其协方差阵$VAR(X)\geqslant0$,$\det VAR(X)=0$的充要条件为:$\exists l_0\in R^m$且$l_0\neq0$,使得

$$\mathrm{Var}(l_0'X)=0.$$

证 显然,从略.

定理3.1 设分布族 $\mathscr{P}=\{P_\theta,\ \theta\in\Theta\}$ 的密度 $\dfrac{dP_\theta}{d\mu}=f(x,\ \theta)$ 满足正则条件 i*)～v*)．而且

$$v(\theta)=\left(E_\theta\left\{(f(x,\ \theta))^{-2}\frac{\partial^i}{\partial\theta^i}f(x,\ \theta)\frac{\partial^j}{\partial\theta^j}f(x,\ \theta)\right\}\right)>0$$

$$(\text{当 }\theta\in\Theta). \tag{3.1}$$

则

$$\mathrm{Var}_\theta\{\hat{g}(X)\}\geqslant D^\tau(\theta)V^{-1}(\theta)D(\theta)\quad(\text{当 }\theta\in\Theta). \tag{3.2}$$

此处

$$D(\theta)=\left(\frac{\partial g(\theta)}{\partial\theta},\ \cdots,\ \frac{\partial^k g(\theta)}{\partial\theta^k}\right)^\tau. \tag{3.3}$$

证 记

$$Z_i(x,\ \theta)=(f(x,\ \theta))^{-1}\frac{\partial^i}{\partial\theta^i}f(x,\ \theta)\ (i=1,\ 2,\ \cdots,\ k). \tag{3.4}$$

$$Z(x,\ \theta)=(Z_1(x,\ \theta),\ \cdots,\ Z_k(x,\ \theta))^\tau. \tag{3.5}$$

由正则条件 iii*)～v*) 知 $E_\theta Z(x,\ \theta)=0$, $VAR_\theta\{Z(x,\ \theta)\}=V(\theta)$ 有限, $\mathrm{cov}(\hat{g},\ Z)=D(\theta)$．所以在 $V(\theta)>0$ 时，我们有

$$0\leqslant VAR\begin{pmatrix}\hat{g}\\ Z\end{pmatrix}=\begin{pmatrix}VAR(\hat{g}) & D^\tau(\theta)\\ D(\theta) & V(\theta)\end{pmatrix},$$

$$0\leqslant\begin{pmatrix}1 & -D^\tau(\theta)V^{-1}(\theta)\\ 0 & I_k\end{pmatrix}\begin{pmatrix}\mathrm{Var}_\theta(\hat{g}) & D^\tau(\theta)\\ D(\theta) & V(\theta)\end{pmatrix}\begin{pmatrix}1 & 0\\ -V^{-1}(\theta)D(\theta) & I_k\end{pmatrix}$$

$$=\begin{pmatrix}\mathrm{Var}\{\hat{g}\}-D^\tau(\theta)V^{-1}(\theta)D(\theta) & 0\\ 0 & V(\theta)\end{pmatrix}. \tag{3.6}$$

故

$$\mathrm{Var}_\theta(\hat{g})\geqslant D^\tau(\theta)V^{-1}(\theta)D(\theta)\quad(\text{当 }\theta\in\Theta). \tag{3.7}$$

定理3.2 在定理 3.1 的条件下，(3.2) 等号成立的充要条件为：存在 $c(\theta)\in R^k$，且 $c(\theta)\neq0$，使

$$\hat{g}(x)=c^\tau(\theta)Z(x,\ \theta)+g(\theta),\quad a.s.P_\theta. \tag{3.8}$$

证 根据引理 3.1 知，(3.2) 等号成立 $\Leftrightarrow\exists l(\theta)=\begin{pmatrix}l_0(\theta)\\ l_1(\theta)\end{pmatrix}\neq0$，

使

$$\mathrm{Var}_\theta[l_0(\theta)\hat{g}+l_1^\tau(\theta)Z(x,\ \theta)]=0. \tag{3.9}$$

此时必有 $l_0(\theta)\neq 0$, 否则, $\mathrm{Var}[l_1^\tau(\theta)Z(x,\ \theta)]=0$, 这与 $V(\theta)>0$ 的假设矛盾. 故存在 $c(\theta)\neq 0$, 使

$$\hat{g}=c^\tau(\theta)Z(x,\ \theta)+g(\theta),\quad a.s.P_\theta. \tag{3.10}$$

反之, 显然.

定理 3.3 在定理 3.1 的条件成立下, Bh 不等式的下界随 k 增大而非降.

证 令

$$D_k(\theta)=\left(\frac{\partial g}{\partial\theta},\ \cdots,\ \frac{\partial^{k-1}g}{\partial\theta^{k-1}}\ \frac{\partial^k g}{\partial\theta^k}\right)^\tau=\begin{pmatrix}D_{k-1}(\theta)\\ \dfrac{\partial^k g}{\partial\theta^k}\end{pmatrix};$$

$$V(\theta)\triangleq\begin{pmatrix}V_{11}(\theta) & V_{12}(\theta)\\ V_{21}(\theta) & V_{22}(\theta)\end{pmatrix},$$

此处 $V_{11}(\theta)$ 为 $(k-1)\times(k-1)$ 方阵;

$$\begin{aligned}&V(\theta)\left(V^{-1}(\theta)-\begin{pmatrix}V_{11}^{-1}(\theta) & 0\\ 0 & 0\end{pmatrix}\right)V(\theta)\\ &=V(\theta)-\begin{pmatrix}V_{11}(\theta) & V_{12}(\theta)\\ V_{21}(\theta) & V_{21}(\theta)V_{11}^{-1}(\theta)V_{12}(\theta)\end{pmatrix}\\ &=\begin{pmatrix}0 & 0\\ 0 & V_{22}(\theta)-V_{21}(\theta)V_{11}^{-1}(\theta)V_{12}(\theta)\end{pmatrix}\geqslant 0.\end{aligned} \tag{3.11}$$

从而

$$V^{-1}(\theta)\geqslant\begin{pmatrix}V_{11}^{-1}(\theta) & 0\\ 0 & 0\end{pmatrix},$$

$$D_k^\tau(\theta)V^{-1}(\theta)D_k(\theta)\geqslant D_{k-1}^\tau(\theta)V_{11}^{-1}(\theta)D_{k-1}(\theta). \tag{3.12}$$

从而定理得证. 由于 $k=1$ 时, Bh 下界就是 C-R 下界, 故从定理 3.3 得出: Bh 的下界比 C-R 下界精确.

定理 3.4(Fend)[4] 设 $\mathscr{P}$ 是如下的指数族:

$$\frac{dP_\theta}{d\mu}=f(x,\ \theta)=h(x)e^{t(x)\psi_1(\theta)+\psi_2(\theta)}\quad(\theta\in\Theta), \tag{3.13}$$

Θ 是开区间, $\psi_1(\theta)$、$\psi_2(\theta)$ 在 Θ 上有有限 k 阶导数, 且 $\psi'(\theta)\neq 0$ (当 $\theta\in\Theta$). 如果 $g(\theta)$ 可估, 满足正则条件 v*). 由 (3.1) 式所定

义的 $V(\theta)>0$, 那么 $g(\theta)$ 的无偏估计 $\hat{g}(x)$ 在 θ 处达到 k 阶 Bh 下界而达不到 $k-1$ 阶下界的充要条件是: $\hat{g}(x)$ 是 $t(x)$ 的 k 阶多项式 $(a.s.P_\theta)$.

证　必要性: 沿用定理 3.1 的记号. 容易计算出

$$
\begin{aligned}
Z_i(x,\ \theta) &= (f(x,\ \theta))^{-1}\frac{\partial^i}{\partial\theta^i}f(x,\ \theta) \\
&= b_{i1}(\theta)t(x)+b_{i2}t^2(x)+\cdots+b_{ii}(\theta)t^i(x)+a_i(\theta),
\end{aligned}
\tag{3.14}
$$

其中 $b_{ii}(\theta)=(\psi_1'(\theta))^i\neq 0\quad(i=1,\ 2,\ \cdots,\ k)$. 记

$$
B(\theta)=\begin{pmatrix} b_{11}(\theta) & 0 & \cdots & \cdots & 0 \\ b_{21}(\theta) & b_{22}(\theta) & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ b_{k1}(\theta) & \cdots & \cdots & b_{kk-1} & b_{kk}(\theta) \end{pmatrix},
$$

$$
T(x)=\begin{pmatrix} t(x) \\ t^2(x) \\ \vdots \\ t^k(x) \end{pmatrix},
$$

$$
a(\theta)=\begin{pmatrix} a_1(\theta) \\ \vdots \\ a_k(\theta) \end{pmatrix}.
$$

根据定理 3.2*), $g(\theta)$ 的无偏估计在 θ 处其方差达到 Bh k 阶下界而达不到 $k-1$ 阶, 必存在 $c(\theta)=[c_1(\theta),\ \cdots,\ c_k(\theta)]^\tau$, 而且 $c_k(\theta)\neq 0$ 使得

$$
\begin{aligned}
\hat{g}(x)-g(\theta) &= c^\tau(\theta)Z(x,\ \theta) \\
&= c^\tau(\theta)B(\theta)T(x)+c^\tau(\theta)a(\theta),\quad a.s.\ P_\theta.
\end{aligned}
\tag{3.15}
$$

注意到 $c_k(\theta)\neq 0$, $b_{kk}(\theta)\neq 0$, 而(3.15)右端 $t^k(x)$ 的系数刚好是 $c_k(\theta)b_{kk}(\theta)$. 故必要性得证.

充分性: 由(3.14)式知

$$
T(x)=B^{-1}(\theta)(Z(x,\ \theta)-a(\theta)),
$$

) 容易验证(3.13)的 $f(x,\theta)$ 满足正则条件 i)～iv*).

而 $B(\theta)$ 为下三角阵，其逆阵 $F(\theta)$ 也是下三角阵，其第 k 个对角线上的元素刚好是 $b_{kk}^{-1}(\theta)$，如果 $\hat{g}(x)$ 是 $t(x)$ 的 k 阶多项式，即可表为 $d^{\tau}T(x)+e$. 此处 $d=(d_1, \cdots, d_k)^{\tau}(d_k\neq 0)$. 故

$$\begin{aligned}\hat{g}(x)&=d^{\tau}B^{-1}(\theta)(Z(x, \theta)-a(\theta))+e\\&=d^{\tau}F(\theta)Z(x, \theta)+e-d^{\tau}F(\theta)a(\theta), \ a.s.P_{\theta}.\end{aligned} \tag{3.16}$$

而且 $Z_k(x, \theta)$ 的系数 $d_kF_{kk}(\theta)\neq 0$，根据定理 3.2 知 $\hat{g}(x)$ 在 θ 处 $(V(\theta)>0)$ 达到 Bh 的 k 阶下界，而达不到 $k-1$ 阶下界.

例 3.1 设 $X_1, X_2, \cdots, X_n$ 为 n 个 iid. 随机变量，$x_1\sim N(\theta, 1)$, $-\infty<\theta<\infty$. 由第二章公式(4.21)可算出 θ^k 的 UMVUE 为

$$\varphi_k(\bar{x})=\left(-\frac{1}{\sqrt{n}}\right)^k H_k(-\sqrt{n}\bar{x}),$$

这里
$$H_k(x)=(-1)^k e^{\frac{x^2}{2}}\frac{d^k}{dx^k}e^{-\frac{x^2}{2}}$$

是 k 阶多项式. 而 $\bar{X}\sim N\left(\theta, \dfrac{1}{n}\right)$ 属于指数族. 以后将证明 $V(\theta)>0, \theta\in R^1$. 故 θ^k 的无偏估计 $\varphi_k(\bar{x})$ 达到 Bh 的 k 阶下界，但达不到 $k-1$ 阶下界.

读者不难验证：例 1.2 中的 $e^{-\lambda}$ 的 UMVUE 达不到任何 k 阶 Bh 下界. 是否能通过 $k\to\infty$ 而能达到呢？这将是下面要讨论的问题.

二、$V(m)$ 为对角阵的指数族

下面我们考虑标准的单参数指数族

$$\frac{dP_\theta}{d\mu}=f(x, \theta)=e^{\theta t(x)+\psi(\theta)} \quad (\theta\in\Theta, \ \Theta \text{ 为开区间}),$$

记 $m(\theta)=E_\theta(t(X))$, $M=\{m(\theta):\theta\in\Theta\}$.

引理 3.2 在上述标准指数族中，θ 与 $m(\theta)$ 在 Θ 上一一对应，其 $m(\theta)$ 的反函数 $\theta(m)$ 在 M 的任一点 m_0 的某个邻域 V_{m_0} 内可展为 Taylor 级数.

证 由第一章所给出的指数族的性质知

$$m(\theta)=-\frac{d\psi(\theta)}{d\theta},\ m'(\theta)=\mathrm{Var}(t(X))>0\ (当\ \theta\in\Theta). \tag{3.17}$$

故 $m(\theta)\uparrow$ 且与 θ 一一对应，存在反函数 $\theta(m)$，$m\in M$. 对任一 $\theta_0\in\Theta$，存在邻域 $V_{\theta_0}\subset\Theta$，当 $\theta+h\in V_{\theta_0}$ 时，有

$$\begin{aligned} e^{-\psi(\theta_0+h)}&=\int_x e^{(\theta_0+h)t(x)}d\mu\leqslant\int_{t(x)<0}e^{-|h|t(x)+\theta_0 t(x)}d\mu \\ &\quad+\int_{t(x)>0}e^{|h|t(x)+\theta_0 t(x)}d\mu \\ &=\sum_{i=0}^{\infty}\frac{|h|^i}{i!}E_{\theta_0}|t(X)|^k<\infty. \end{aligned}$$

故

$$e^{-\psi(\theta)}=\sum_{i=0}^{\infty}\frac{(\theta-\theta_0)^i}{i!}E_{\theta_0}(t^i(X))\quad(当\ \theta\in V_{\theta_0}). \tag{3.18}$$

因此适当缩小邻域 V_{θ_0}，可使

$$\psi(\theta)=\sum_{i=0}^{\infty}a_i(\theta-\theta_0)^i\ (当\ \theta\in V_{\theta_0}). \tag{3.19}$$

故 $$-\psi'(\theta)=m(\theta)=\sum_{i=1}^{\infty}ia_i(\theta-\theta_0)^i\quad(当\ \theta\in V_{\theta_0}).$$

由于 $a_1=m'(\theta_0)>0$，根据幂级数反演定理，存在 $m(\theta_0)=m_0\in M$ 的邻域 $V_{m_0}\subset M$，使得 $\theta(m)$ 可展为如下级数：

$$\theta(m)=\sum_{i=0}^{\infty}b_i(m-m_0)^i\quad(当\ m\in V_{m_0}), \tag{3.20}$$

其中 $b_i=\frac{1}{i!}\theta^{(i)}(m_0)$. 故引理得证.

根据引理 3.2，我们可用 m 代 θ 作为参数，以下记

$$\tilde{f}(x,m)=f(x,\theta(m))=\exp\{\theta(m)t(x)+\psi(\theta(m))\}\quad(m\in M);$$

$$Z_i(x,m)=(\tilde{f}(x,m))^{-1}\frac{\partial^i}{\partial m^i}\tilde{f}(x,m)\quad(i=1,2,\cdots);$$

$$V_{ij}(m)=E_m(Z_i(x,m)Z_j(x,m))\quad(i,j=1,2,\cdots);$$

$$\begin{aligned} V_k(m)&=VAR\{Z^{(k)}(x,m)\} \\ &=VAR\{Z_1(x,m),\cdots,Z_k(x,m)\}^{\tau}\ (k=1,2,\cdots). \end{aligned}$$

$V_k(m)$称为$\{\mathcal{f}(x, m), m\in M\}$的 k-阶 Fisher 信息阵.

定理 3.5 在上述假设和记号下，任意 k-阶 Fisher 信息阵皆是对角阵的充要条件是

$$\theta'(m)=[Am^2+B+C]^{-1}>0 \quad (当\ m\in M), \tag{3.21}$$

其中 A、B、C 是常数.

证 由引理 3.2 知，$\theta(m)$ 可在任意 $m_0\in M$ 的一个邻域内展为 Taylor 级数；$\psi(\theta)$ 可展为(3.19)式级数. 因此，$\psi(\theta(m))$ 在 m_0 的某邻域内展为 Taylor 级数，再加上 e^x 的性质，就存在 $\delta>0$(与 x 无关)，当 $|h|<\delta$ 时，

$$\begin{aligned}&(\mathcal{f}(x, m_0))^{-1}(\mathcal{f}(x, m_0+h)-\mathcal{f}(x, m_0))\\&=\sum_{k=1}^{\infty}Z_k(x, m)\frac{h^k}{k!}.\end{aligned} \tag{3.22}$$

两边乘 $Z_l(x, m)$，再求数学期望，利用 $E_m(Z(X, m))=0$ 及(3.22)的级数绝对收敛性，就有($|h|<\delta$)

$$\begin{aligned}&\sum_{k=1}^{\infty}V_{kl}(m_0)\frac{h^k}{k!}\\&=E_{m_0}\{Z_l(X, m_0)\mathcal{f}(X, m_0+h)(\mathcal{f}(X, m_0))^{-1}\}\\&=\exp\{\psi(\theta(m_0+h))-\psi(\theta(m_0))\}\\&\quad\times E_{m_0}[Z_l(X, m_0)\exp\{(\theta(m_0+h)-\theta(m_0))t(X)\}].\end{aligned} \tag{3.23}$$

记 $u=\theta(m_0+h)-\theta(m_0)$. 由于对指数族可在积分号下求导，故

$$\begin{aligned}&E_{m_0}[Z_l(X, m_0)\exp\{\theta(m_0+h)-\theta(m_0)t(X)\}]\\&=\int\frac{\partial^l}{\partial m^l}\mathcal{f}(x, m)e^{ux}\,d\mu\,\Big|_{m=m_0}=\frac{\partial^l}{\partial m^l}\int\mathcal{f}(x, m)e^{ux}\,d\mu\,\Big|_{m=m_0}\\&=\frac{d^l}{dm^l}e^{\psi(\theta(m))-\psi(\theta(m)+u)}\Big|_{m=m_0}.\end{aligned}$$

以此代入(3.23)式，就得到

$$\begin{aligned}&\sum_{k=1}^{\infty}V_{kl}(m_0)\frac{h^k}{k!}\\&=e^{\psi(\theta(m_0+h))-\psi(\theta(m_0))}\frac{d^l}{dm^l}e^{\psi(\theta(m))-\psi(\theta(m)+u)}\Big|_{m=m_0}\end{aligned} \tag{3.24}$$

必要性：设 $V_{kl}(m)=\delta_{kl}V_{kk}(m)$，此处

$$\delta_{kl}=\begin{cases}1, \text{ 当 } k=l;\\ 0, \text{ 当 } k\neq l.\end{cases}$$

取 $l=1$，(3.24)式就成为

$$V_{11}(m_0)h=e^{\psi(\theta(m_0+h))-\psi(\theta(m_0))}$$

$$\frac{d}{dm}\exp\{\psi(\theta(m))-\psi(\theta(m)+u)\}\big|_{m=m_0}$$

$$=e^{\psi(\theta(m_0+h))-\psi(\theta(m_0)+u)}[\psi'(\theta(m_0))\theta'(m_0)$$

$$-\psi'(\theta(m_0)+u)\theta'(m_0)]$$

$$=[\psi'(\theta(m_0))-\psi'(\theta(m_0+h))]\theta'(m_0). \qquad (3.25)$$

(3.25)最后一个等式是用 $u=\theta(m_0+h)-\theta(m_0)$ 代入而得到，此处

$$\psi'(\theta)=\frac{d}{d\theta}\psi(\theta);\quad \theta'(m)=\frac{d}{dm}\theta(m).$$

众所周知，$\psi'(\theta)=-m(\theta)$，以此代入(3.25)式，得

$$V_{11}(m_0)h=\theta'(m_0)h \quad (\text{当}|h|<\delta).$$

因此 $$V_{11}(m)=\theta'(m) \quad (\text{当 } m\in M).$$

再取 $l=2$，通过与 $l=1$ 时类似的推演可得

$$V_{22}(m_0)\ \frac{h^2}{2}=(\theta'(m_0))^2h^2$$

$$+[-\theta'(m_0)+h\theta''(m_0)-\psi''(\theta(m_0+h))(\theta'(m_0))^2],$$

也即

$$\psi''(\theta(m_0+h))=\Big\{-\theta'(m_0)+\theta''(m_0)h$$

$$+\Big[(\theta'(m_0))^2-\frac{1}{2}V_{11}(m_0)\Big]h^2\Big\}(\theta'(m_0))^{-2}. \qquad (3.26)$$

根据引理3.2的推导，可在 $|h|<\delta$ 内把 $\psi''(\theta(m_0+h))$ 展为 h 的幂级数 $\sum\limits_{i=0}^{\infty}\frac{1}{i!}r^{(i)}(m_0)h^i$. 此处

$$r^{(i)}(m_0)=\frac{d^i}{dm^i}\psi''(\theta(m)).$$

以 $\sum\limits_{i=0}^{\infty}\frac{1}{i!}r^{(i)}(m_0)h^i$ 代入(3.26)式，再对 $h^i(i=0, 1, 2, \cdots)$ 的系数比较得

$$r(m_0)=-[\theta'(m_0)]^{-1},\quad r'(m_0)=\theta''(m_0)[\theta'(m_0)]^{-2},$$

$$r''(m_0)=\left[(\theta'(m_0))^2-\frac{1}{2}V_{22}(m_0)\right][\theta'(m_0)]^{-2},$$

$$r^{(i)}(m_0)=0\quad(当\ i\geqslant 3).$$

由于 m_0 是 M 中的任意一点，根据 $r^{(i)}(m)=0$(当 $i\geqslant 3$)，$m\in M$ 皆成立. 由此可知

$$r(m)=-[\theta'(m)]^{-1}=-(Am^2+Bm+C)\quad(当\ m\in M).$$

故
$$\theta'(m)=(Am^2+Bm+C)^{-1}.$$

必要性得证.

充分性：这是一个繁琐的分析，以下分为 $A=B=0$；$A=0$, $B\neq 0$; $A>0$, 但 $Am^2+Bm+C=0$ 有两个相等实根; $A>0$, $Am^2+Bm+C=0$ 有两个不等实根; $A<0$, Am^2+Bm+C 有两个不等实根；$A>0$ 无实根等六种情况进行讨论. 由于证明的方法基本相同,因此只详细证明 $A=B=0$ 的情况,其它情况只叙述其结果,证明从略,留作习题.

(1) $A=B=0$,

$$\tilde{f}(x,\ m)=\exp\left\{c^{-1}t(x)-\frac{1}{2}c^{-1}m^2\right\},\ c>0. \tag{3.27}$$

$$t(X)\sim N(m,\ c). \tag{3.28}$$

$$V_{ij}(m)=i!c^{-i}\delta_{ij}\quad(i,\ j=1,\ 2,\ \cdots). \tag{3.29}$$

因 $\theta'(m)=(m'(\theta))^{-1}=(\mathrm{Var}_\theta(X))^{-1}>0$, 故 $c>0$.

由方程 $\theta'(m)=c^{-1}$ 解得 $\theta(m)=c^{-1}m+d$, 而

$$\frac{d}{dm}\psi(\theta(m))=-m\theta'(m)=-c^{-1}m$$

解得 $\psi(\theta(m))=-\frac{1}{2}c^{-1}m^2+d_2=-\frac{1}{2}c(\theta(m)-d)^2+d_1.$

而 $t(x)$ 的特征函数为

$$f(t)=\exp\{\psi(\theta)-\psi(\theta+it)\}=\exp\left\{-\frac{ct^2}{2}+imt\right\}.$$

这说明 $t(X)\sim N(m,\ c)$.

由于 $\psi(\theta(m))-\psi(\theta(m)+u)=mu+\frac{cu^2}{2}$, 将此式代入(3.24), 可得当 $|h|<\delta$ 时,

$$\sum_{k=1}^{\infty} V_{kl}(m) \frac{h^k}{k!} = e^{\psi(\theta(m+h))-\psi(\theta(m))} \frac{d^l}{dm^l} e^{mu+\frac{cu^2}{2}}$$

$$= e^{-c^{-1}h-\frac{c^{-1}h^2}{2}} u^l e^{mu+\frac{cu^2}{2}}.$$

我们用原来 u 的定义为 $u=\theta(m+h)-\theta(m)=c^{-1}h$ 代入上式右端，即得

$$\sum_{k=1}^{\infty} V_{kl}(m) \frac{h^k}{k!} = c^{-l}h^l.$$

由此即得 $\qquad V_{kl}=k!c^{-l}\delta_{kl} \quad (k,\ l=1,\ 2,\ \cdots).$

总结其证明步骤为：

i) 利用 $\theta'(m)=(Am^2+Bm+C)^{-1}$ 及 $\frac{d}{dm}\psi(\theta(m))=m\theta'(m)$ 可解得到 $\theta(m)$ 及 $\psi(\theta(m))$ 的一般表达式．从而可得 $\tilde{f}(x,\ m)$ 的一般表达式．

ii) 利用 $e^{\psi(\theta(m))-\psi(\theta(m)+iu)} \triangleq \varphi(u)$（这是 $t(X)$ 的特征函数）可找出 $t(X)$ 的具体分布函数．

iii) 以 $\psi(\theta(m))-\psi(\theta(m)+u)$ 代入(3.24)式的右端，取微分（u 作为与 m 无关的参变量），然后利用 $u=\theta(m_0+h)-\theta(m_0)$ 再展为 h 的幂级数，与等式左端对比，即得 $V_{kl}(m_0)$ 的表达式．

根据以上步骤，就可推得以下结论：

(2) $A=0,\ B\neq 0,\ Bm+C>0$ （当 $m\in M$ 时）．

$$\tilde{f}(x,\ m)=\exp\{B^{-1}(t(x)+B^{-1}C)\log(Bm+C) - B^{-2}(Bm+C)\}. \tag{3.30}$$

$t(X)$ 的特征函数为

$$\varphi(u)=e^{-B^{-2}(Bm+C)(1-e^{iBu})-iB^{-1}Cu}. \tag{3.31}$$

$B^{-1}t(X)+B^{-1}C\sim$Poisson 分布 $\mathscr{P}(B^{-1}m+c)$．此时必有 $B=0$ 或 $C=0$．

$$V_{ij}(m)=i!\,(Bm+C)^{-i}\delta_{ij} \quad (i,\ j=1,\ 2,\ \cdots). \tag{3.32}$$

(3) $A>0,\ Am^2+Bm+C=A(m-\alpha)^2,\ m>\alpha$ （或 $m<\alpha$）．

$$\tilde{f}(x,\ m)=\exp\{-A^{-1}(m-\alpha)^{-1}(t(x)-\alpha)-A^{-1}\log|m-\alpha|\}. \tag{3.33}$$

$t(X)$ 的特征函数为

$$\varphi(u)=e^{i\alpha u}(1-iA(m-\alpha)u)^{-A^{-1}}. \tag{3.34}$$

$\pm t(X)-\alpha$ 服从 Γ 分布族 $G(|\theta|, A^{-1})$.

$$V_{ij}(m)=i!\left(\prod_{k=0}^{i-1}(A^{-1}+k)\right)|m-\alpha|^{-2i}\delta_{ij} \quad (ij=1, 2, \cdots). \tag{3.35}$$

(4) $A>0$, $Am^2+Bm+C=A(m-\alpha)(m-\beta)$, $\alpha<\beta$, $m>\beta$ (或 $m<\alpha$).

$$\mathfrak{f}(x, m)=\exp\left\{rt(X)\log\frac{m-\beta}{m-\alpha}-r\log\frac{|m-\beta|^{\beta}}{|m-\alpha|^{\alpha}}\right\},$$
$$r=(\beta-\alpha)^{-1}A^{-1}. \tag{3.36}$$

$t(X)$ 的特征函数为

$$\varphi(u)=e^{iau}P^{\frac{1}{A}}(1-qe^{ir^{-1}u})^{-1/A}. \tag{3.37}$$

此处 $p=1-q$, $0<p<1$. 在 $m>\beta>\alpha$ 时,

$$a=\beta,\ p=1-\frac{m-\beta}{m-\alpha};$$

在 $m<\alpha<\beta$ 时,

$$a=\alpha,\ p=1-\frac{\alpha-m}{\beta-m}.$$

$rt(z)-a$ 的分布为

$$P(rt(X)-a=k)=\binom{-A^{-1}}{k}q^kp^{A^{-1}}(-1)^k.$$

在 A^{-1} 为正整数时,刚好是逆二项分布.

$$V_{ij}(m)=i!\left(\prod_{k=0}^{i-1}(r\beta-r\alpha+k)\right)((\alpha-m)(\beta-m))^{-i}$$
$$(i, j=1, 2, \cdots). \tag{3.38}$$

(5) $A<0$, $Am^2+Bm+C=A(m-\alpha)(m-\beta)$, $\alpha<m<\beta$.

$$\mathfrak{f}(x, m)=\exp\left\{rt(x)\log\frac{m-\alpha}{\beta-m}-r\log\frac{(m-\alpha)^{\alpha}}{(\beta-m)^{\beta}}\right\},$$
$$r=-A^{-1}(\beta-\alpha)^{-1}. \tag{3.39}$$

$t(X)$的特征函数为

$$\varphi(u)=e^{i\alpha u}(p+qe^{ir^{-1}u})^{-A^{-1}}. \tag{3.40}$$

此处 $p=1-q,\ 0<p<1,\ p=\left(1+\frac{m-\alpha}{\beta-m}\right)^{-1}$. 在 $-A^{-1}$ 为自然数时，$rt(X)-\alpha\sim$ 二项分布 $B(p,\ -A^{-1})$.

$$V_{ij}(m)=i!\left(\prod_{k=0}^{i-1}(r\beta-r\alpha-k)\right)[(m-\alpha)(\beta-m)]^{-i}\delta_{ij}$$

$$(i,\ j=1,\ 2,\ \cdots). \tag{3.41}$$

此式表明 $r(\beta-\alpha)=-A^{-1}$ 只能是非负整数方能有 $V_{ii}(m)\geqslant 0$，否则，会有 $V_{ii}(m)<0$ 出现. 这与 $V_{ii}(m)$ 的含意不符. 故这种情况仅能是二项分布.

(6) $A>0,\ Am^2+Bm+C\,(m\in R^1)$ 无实根.

$$\tilde{f}(x,\ m)=\exp\left\{r\left(t(x)-\frac{\alpha}{\beta}\right)\operatorname{arc\,tg}(\beta m-\alpha)\right.$$

$$\left.-\frac{\gamma}{2\beta}\log(1+(\beta m-\alpha)^2)\right\}. \tag{3.42}$$

此处 $r=2(4AC-B^2)^{-\frac{1}{2}},\ \beta=rA>0,\ \alpha=-\frac{rB}{2}$.

$t(X)$ 的特征函数为

$$g(u)=(\cos^2 r^{-1}\theta)^{\frac{r}{2\beta}}(\cos^2(r^{-1}(\theta+iu)))^{-\frac{r}{2\beta}}e^{i\frac{\alpha}{\beta}u}. \tag{3.43}$$

其对应的分布在[18]有详尽的讨论.

$$r_{ij}(m)=i!\beta^{2i}\left(\prod_{k=0}^{i-1}\left(\frac{\gamma}{\beta}+k\right)\right)(1+(\beta m-\alpha)^2)^{-i}\delta_{ij}$$

$$(i,\ j=1,\ 2,\ \cdots). \tag{3.44}$$

以上六种情况 $V_k(m)\ (k=1,\ 2,\ \cdots)$ 皆是对角阵，从而定理得证.

不难看到满足 $\theta'(m)=(Am^2+Bm+C)^{-1}>0$ 的指数型分布族包含了常见的几种类型分布，因而我们称它为常用指数型分布族.

定理 3.6 如果 $\{\tilde{f}(x,\ m)=f(x,\ \theta(m)),\ m\in M\}$ 为常用指数族，$g(m)$ 是 M 上的可估函数，那么 $g(m)$ 有有限方差的 UMVUE $\hat{g}$ 存在的充要条件是：当 $m\in M$，有

$$\sum_{k=1}^{\infty}(g^{(k)}(m))^2V_{kk}^{-1}(m)<\infty. \tag{3.45}$$

此处规定 $V_{kk}^{-1}(m)=0$(当 $V_{kk}(m)=0$ 时)．而且

$$\mathrm{Var}_m(\hat{g})=\sum_{k=1}^{\infty}(g^{(k)}(m))^2V_{kk}^{-1}(m)\quad(\text{当 } m\in M). \quad (3.46)$$

证 充分性：设 (3.45) 成立，在 M 中任取一点 m_0，由于 $g(m)$ 可估，令 $\tilde{g}(x)$ 为 $g(m)$ 的无偏估计．由于 $t(x)$ 为充分统计量，故

$$\hat{g}(t(x))=E(\hat{g}(X)\,|\,t(x))$$

也是 $g(m)$ 的无偏估计．根据指数族分布性质：

$$g(m)-g(m_0)=\sum_{i=1}^{\infty}g^{(i)}(m_0)\frac{(m-m_0)^i}{i!}\quad(\text{当 } m\in M).$$

记 $\quad \zeta_0=g(m_0),\ \zeta_i=g^{(i)}(m_0)V_{ii}^{-\frac{1}{2}}(m_0)\quad(i=1,\ 2,\ \cdots)$,

由假设 $\sum_{i=1}^{\infty}\zeta_i^2<\infty$ 以及 $\{1,\ Z_i(x,\ m_0)V_{ii}^{-\frac{1}{2}}(m_0),\ i=1,\ 2,\ \cdots\}$ 是 $L_2(\tilde{f}(x,\ m_0)d\mu)$ 的一组完全正交基．存在

$$\hat{g}_0(t(x))\in L_2(\tilde{f}(x,\ m_0)d\mu)$$

且

$$E_{m_0}(\hat{g}_0(t(X))-\sum_{i=0}^{n}\zeta_iZ_i(X,\ m_0)V_{ii}^{-\frac{1}{2}}(m_0))^2\to 0$$

$$(\text{当 } n\to\infty).$$

$$E_{m_0}(\hat{g}_0(t(X))Z_i(X,\ m_0))=\zeta_i\ (i=1,\ 2,\ \cdots),$$

$$E_{m_0}[\hat{g}_0(t(X))]^2=\sum_{i=0}^{\infty}\zeta_i^2<\infty.$$

又根据引理 3.2, 存在 $\delta>0$, 当 $|m-m_0|<\delta$ 时,

$$\begin{aligned}\tilde{f}(x,\ m)&=\sum_{i=0}^{\infty}\frac{\partial^i\tilde{f}(x,\ m_0)}{\partial m_0^i}(m-m_0)^i/i!\\&=\sum_{i=0}^{\infty}Z_i(x,\ m_0)\tilde{f}(x,\ m_0)(m-m_0)^i/i!\,(\text{当 } x\in R^1).\end{aligned}$$

因为

$$\begin{aligned}&\int|\hat{g}_0(t(x))Z_i(x,\ m_0)|\tilde{f}(x,\ m_0)d\mu\\&\leqslant(E_{m_0}[\hat{g}_0(t(X))]^2)^{\frac{1}{2}}[E_{m_0}\{Z_i^2(X,\ m_0)\}]^{\frac{1}{2}}\\&=\Big(\sum_{i=0}^{\infty}\zeta_i^2\Big)^{\frac{1}{2}}\sqrt{V_{ii}(m_0)}.\end{aligned}$$

根据定理 3.5 的充分性证明中关于 $V_{ii}(m_0)$ 的表达式知道，存在 $k_1(m_0)$，使得

$$V_{ii}(m_0)\leqslant(i!)^2k_1^i(m_0)\quad(i=1,\ 2,\ \cdots).$$

故当 $|m-m_0|<\min(\delta,\ k_1^{-1/2}(m_0))\triangleq\delta_0$ 时，

$$\sum_{i=0}^{\infty}\int|Z_i(x,\ m_0)\hat{g}_0(t(x))|f(x,\ m_0)d\mu\ \frac{1}{i!}<\infty.$$

因此当 $|m-m_0|<\delta_0$ 时，

$$\begin{aligned}
&E_m[\hat{g}_0(t(X))]\\
&=\int\hat{g}_0(t(x))\sum_{i=0}^{\infty}Z_i(x,\ m_0)f(x,\ m_0)(m-m_0)^i/i!\\
&=\sum_{i=0}^{\infty}\int\hat{g}_0(t(x))Z_i(x,\ m_0)f(x,m_0)\,d\mu\cdot\frac{(m-m_0)^i}{i!}\\
&=\sum_{i=0}^{\infty}\zeta_iV_{ii}^{\frac{1}{2}}(m_0)\ \frac{(m-m_0)^i}{i!}=\sum_{i=0}^{\infty}g^{(i)}(m_0)\ \frac{(m-m_0)^i}{i!}\\
&=g(m).
\end{aligned}$$

由于 $t(X)$ 是 θ 在 Θ 上的任意区间上（或 m 在 M 上任意区间上）为完全充分统计量，从而

$$\hat{g}_0(t(X))=\hat{g}(t(x))\quad a.s.\mu.$$

故

$$\begin{aligned}
\mathrm{Var}_{m_0}(\hat{g}(t(X)))&=\mathrm{Var}_{m_0}(\hat{}_0(t(X)))\\
&=\sum_{k=1}^{\infty}(g^{(i)}(m_0))^2V_{kk}^{-1}(m_0)<\infty.
\end{aligned}$$

由于 m_0 的任意性，充分性得证。

必要性：因为 $g(m)$ 可估，存在有限方差的无偏估计 $\tilde{g}(x)$。利用 B-R-L-S 定理，$E(\tilde{g}(X)|t(x))=\hat{g}(t(x))$，即是 UMVUE。再由 Bh 不等式及 $g^{(\nu)}(m)\ (\nu=1,\ 2,\ \cdots)$ 存在，故有

$$\infty>\mathrm{Var}_m(\hat{g}(t(X)))\geqslant\sum_{\nu=1}^{k}(g^{(\nu)}(m))^2V_{\nu\nu}^{-1}(m)$$

$$(\text{当 } m\in M,\ k=1,\ 2,\ \cdots).$$

故

$$\sum_{\nu=1}^{\infty}(g^{(\nu)}(m))^2V_{\nu\nu}^{-1}(m)<\infty.$$

再利用充分性的结论，即得

$$\mathrm{Var}_m(\hat{g}(t(X)))=\sum_{\nu=1}^{\infty}(g^{(\nu)}(m))^2V_{\nu\nu}^{-1}(m)\quad(当\ m\in M).$$

至此定理证完.

前面虽曾指出了本章的例1.2中的 $g(\lambda)=e^{-\lambda}$ 的无偏估计方差达不到任意阶 Bh 下界. 但由于 Poisson 分布属于常用的指数族分布,故

$$\mathrm{Var}\left(\left(1-\frac{1}{n}\right)^{\sum_1^n x_i}\right)=\sum_{k=1}^{\infty}e^{-2\lambda}V_{kk}^{-1}(\lambda)=e^{-2\lambda}(e^{\frac{\lambda}{n}}-1)$$

达到了 $k\to\infty$ 时的 Bh 下界.

例 3.2 设 $X\sim\Gamma$ 分布,其密度函数为

$$f(x,\ \theta)=(\theta\Gamma(\alpha))^{-1}(x\theta^{-1})^{\alpha-1}e^{-\frac{x}{\theta}}\quad(\theta>0,\ \alpha>0,\ x>0),$$

$EX=\alpha\theta\triangleq m,\ g(m)=e^{-\frac{m}{\alpha}}$; 其 UMVUE 为

$$\hat{g}(x)=\sum_{k=0}^{\infty}\frac{\Gamma(\alpha)}{\Gamma(k+\alpha)}\frac{(-1)^k}{k!}x^k;$$

$$V_{kk}(m)=k!\prod_{i=0}^{k-1}(i+\alpha)m^{-2k}=k!\Gamma(k+\alpha)\Gamma^{-1}(\alpha)m^{-2k}$$

$$(k=1,\ 2,\ \cdots);$$

$$\mathrm{Var}_m(\hat{g}(X))=\sum_{k=1}^{\infty}\frac{\Gamma(\alpha)}{\Gamma(k+\alpha)}\cdot\frac{\theta^{2k}}{k!}e^{-2\theta}.$$

这些公式由(3.35)和定理3.6的(3.46)式可得到,但另一方面,又可直接计算 $\hat{g}(x)$ 的方差 $E(\hat{g}^2(X))-(Eg(X))^2$ 而得到恒等式

$$\sum_{n=0}^{\infty}\sum_{k=0}^{n}(-1)^n\frac{\Gamma(n+\alpha)\Gamma(\alpha)}{\Gamma(k+\alpha)\Gamma(n-k+\alpha)}\cdot\frac{\theta^n}{k!(n-k)!}-e^{-2\theta}$$
$$=\sum_{k=1}^{\infty}\frac{\Gamma(\alpha)}{\Gamma(k+\alpha)}\cdot\frac{\theta^{2k}e^{-2\theta}}{k!}.\tag{3.47}$$

§4 多参数的C-R型不等式

一、C-R 不 等 式

此节是把单参数的 C-R 不等式与 Bh 不等式的某些结果推广到多参数的情况. 我们仍然采用本章以往采用过的符号与相关

的假设．现在先讨论 C-R 不等式．仍然假设下列正则条件成立：

i**) Θ 是 R^p 上非空开区域；

ii**) 关于 μ 的密度 $\dfrac{dP_\theta}{d\mu}=f(x,\ \theta)\ (\theta\in\Theta)$ 有共同的支撑，不妨设 $f(x,\ \theta)>0$ (当 $x\in X,\ \theta\in\Theta$)；

iii**) $\displaystyle\int\frac{\partial}{\partial\theta}f(x,\ \theta)d\mu=0$；

iv**) $\displaystyle\int f^{-1}(x,\ \theta)\left|\frac{\partial}{\partial\theta}f(x,\ \theta)\right|^2d\mu<\infty$ (当 $\theta\in\Theta$)；

v**) Fisher 信息阵

$$I(\theta)=\int f^{-1}(x,\ \theta)\left(\frac{\partial}{\partial\theta}f(x,\ \theta)\right)\left(\frac{\partial}{\partial\theta}f(x,\ \theta)\right)^\tau d\mu>0$$

$$(\text{当 }\theta\in\Theta).$$

引理 4.1 设 n 阶方阵 $A>0,\ B\geqslant0,\ A-B\geqslant0$．则

$$A=B\Leftrightarrow\det A=\det B.$$

证 因 $A>0,\ B\geqslant0$．所以存在非奇异方阵 G，使

$$A=G'G;\quad B=G'\Lambda G.$$

此处 $\Lambda=\mathrm{diag}(\lambda_1,\ \lambda_2,\ \cdots,\ \lambda_n)$，且 $\lambda_i\geqslant0\quad(i=1,\ 2,\ \cdots,\ n)$．

又 $$0\leqslant A-B=G'(I-\Lambda)G\Rightarrow I-\Lambda\geqslant0.$$

故 $0\leqslant\lambda_i\leqslant1\ (i=1,\ 2,\ \cdots,\ n)$，从而可知

$$A=B\Leftrightarrow I=\Lambda\Leftrightarrow\det A=\det B.$$

定理 4.1 设分布族 $\mathscr{P}$ 满足正则条件 i**)～v**)．

$$G(\theta)=(g_1(\theta),\ \cdots,\ g_k(\theta))',\ k\leqslant p$$

可估，且是 Θ-R^p 连续可微映象．记

$$D(\theta)=\left(\frac{\partial g_i}{\partial\theta_j}\right)_{k\times p},\ \hat{G}(x)=(\hat{g}_1(x),\ \cdots,\ \hat{g}_k(x))',$$

且 $E_\theta\hat{G}(X)=G(\theta)$，$|\mathrm{VAR}_\theta\hat{G}(X)|<\infty$ (当 $\theta\in\Theta$)．$\hat{G}(x)$ 还满足正则条件

vi**) $$\int\hat{G}(x)\left(\frac{\partial}{\partial\theta}f(x,\ \theta)\right)'d\mu=\left(\frac{\partial}{\partial\theta}\int\hat{G}(x)f(x,\ \theta)d\mu\right)$$

$$=D(\theta)\quad(\text{当 }\theta\in\Theta),$$

那么

$$\mathrm{VAR}_\theta(\hat{G}) \geqslant D(\theta) I^{-1}(\theta) D'(\theta); \tag{4.1}$$

$$\det \mathrm{VAR}_\theta(\hat{G}) \geqslant \det(D(\theta) I^{-1}(\theta) D'(\theta)) \quad (当\ \theta \in \Theta). \tag{4.2}$$

若 $p \leqslant k$, $D(\theta)$ 在 $\theta=\theta_0 \in \Theta$ 满秩，则(4.1)、(4.2)在 $\theta=\theta_0$ 处等号成立的充要条件为：存在 $k \times p$ 阵 $B(\theta_0)$，使

$$\hat{G}-G=B(\theta_0) Z(x, \theta_0) \triangleq B(\theta_0) \frac{\partial}{\partial \theta_0} \log f(x, \theta_0), \quad a.s. P_\theta. \tag{4.3}$$

其中 $\operatorname{rank} B(\theta_0)=\operatorname{rank}(VAR_{\theta_0}(\hat{G}))$.

证 由 iii**) 知 $E_\theta Z(x, \theta_0)=0$, 由 iv**) 及 v**) 知

$$|I(\theta)|<\infty, \ I(\theta)>0 \quad (当\ \theta \in \Theta).$$

由正则条件 vi**) 有

$$\operatorname{cov}_\theta(\hat{G}, Z(x, \theta))=E_\theta\left\{\hat{G}(X)\left(\frac{\partial}{\partial \theta} \log f(X, \theta)\right)'\right\}$$

$$=\frac{\partial}{\partial \theta}(E_\theta \hat{G}(X))=D(\theta) \quad (当\ \theta \in \Theta). \tag{4.4}$$

从而

$$0 \leqslant \mathrm{VAR}_\theta\begin{pmatrix}\hat{G} \\ Z(x, \theta)\end{pmatrix}=\begin{pmatrix}\mathrm{VAR}_\theta \hat{G} & \operatorname{cov}_\theta(\hat{G}, Z) \\ \operatorname{cov}_\theta(Z, \hat{G}) & \mathrm{VAR}_\theta Z\end{pmatrix}$$

$$=\begin{pmatrix}\mathrm{VAR}_\theta \hat{G} & D(\theta) \\ D'(\theta) & I(\theta)\end{pmatrix}. \tag{4.5}$$

根据矩阵公式(3.6)，当 $I(\theta)>0$ 时，就推得

$$\mathrm{VAR}_\theta \hat{G}-D(\theta) I^{-1}(\theta) D'(\theta) \geqslant 0.$$

从而(4.1)、(4.2)式成立.

现证定理的第二部分：因

$$I(\theta_0)>0, \ \operatorname{Rank} D(\theta_0)=k,$$

故存在方阵 $F(\theta_0)$，使

$$I^{-1}(\theta_0)=F(\theta_0)(F(\theta_0))',$$

$$k=\operatorname{Rank} DF=\operatorname{Rank} DFF'D'=\operatorname{Rank} DI^{-1}(\theta_0) D'.$$

以致 $\quad D(\theta_0) I^{-1}(\theta_0)(D(\theta_0))'>0, \ \mathrm{VAR}_{\theta_0}(\hat{G})>0.$

由引理 4.1 知

$$\mathrm{VAR}_{\theta_0} \hat{G}=D(\theta_0) I^{-1}(\theta_0)(D(\theta_0))' \Leftrightarrow \det \mathrm{VAR}_{\theta_0} \hat{G}$$

$$=\det(D(\theta_0) I^{-1}(\theta_0)(D(\theta_0))'),$$

必要性：对任意 $k\times p$ 阵 B，有

$$\operatorname{cov}_{\theta_0}(\hat{G},\ BZ(x,\ \theta_0))=E_{\theta_0}(\hat{G}-G)Z'(x,\ \theta)B')$$
$$=D(\theta_0)B'; \tag{4.6}$$
$$\operatorname{cov}_{\theta_0}(BZ(x,\ \theta_0),\ \hat{G})=BD(\theta_0)'; \tag{4.7}$$
$$\mathrm{VAR}_{\theta_0}(\hat{G}-BZ(x,\ \theta_0))$$
$$=\mathrm{VAR}_{\theta_0}(\hat{G})-D(\theta_0)B'-B'D'(\theta_0)+\mathrm{VAR}_{\theta_0}(BZ(x,\ \theta_0))$$
$$=\mathrm{VAR}_{\theta_0}(\hat{G})+BI(\theta_0)B'-D(\theta_0)B'-BD'(\theta_0). \tag{4.8}$$

若(4.1)或(4.2)在 $\theta=\theta_0$ 处等号成立，取 $B=D(\theta_0)I^{-1}(\theta_0)$，将它代入(4.8)式，即得

$$\mathrm{VAR}_{\theta_0}(\hat{G}-BZ(x,\ \theta_0))=0,$$

故 $\qquad \hat{G}-G(\theta_0)=I^{-1}(\theta_0)D'(\theta_0)Z(x,\ \theta_0),\ a.s.P_{\theta_0}.$

因 $\operatorname{Rank} D(\theta_0)=k$，故 $\operatorname{Rank} D(\theta_0)I^{-1}(\theta_0)=k$. 必要性得证.

充分性：若有 $B(\theta_0)$ 使 $\hat{G}-G(\theta)=B(\theta_0)Z(x,\ \theta_0)$，$a\ s.P_{\theta_0}$，故

$$E_{\theta_0}(\hat{G}-G(\theta_0))Z'(x,\ \theta_0)=D(\theta_0)=B(\theta_0)I(\theta_0).$$

也即 $\qquad B(\theta_0)=D(\theta_0)I^{-1}(\theta_0).$

将它代入(4.8)式，得

$$0=\mathrm{VAR}_{\theta_0}(\hat{G})-D(\theta_0)I^{-1}(\theta_0)(D(\theta_0))'.$$

这说明(4.1)式在 $\theta=\theta_0$ 处等号成立. 因而(4.2)也成立. 至此，定理证毕.

定理 4.2 若分布族 $\mathscr{P}$ 满足正则条件 i**)～v**)，$\hat{\theta}(x)$ 是 θ 的估计，且

$$E_\theta(\hat{\theta}(X))=\theta+b(\theta)\triangleq G(\theta),\quad \tilde{D}(\theta)\triangleq\left(\frac{\partial b(\theta)}{\partial\theta}\right),$$

又设 $\hat{\theta}(x)$ 满足正则条件 vi**)，则对任意的 $\zeta\neq 0$ 的 p 维向量，皆有

$$\zeta'\mathrm{VAR}_\theta(\hat{\theta}(X))\zeta$$
$$\geqslant\zeta'I^{-1}(\theta)\zeta[1+\zeta'\tilde{D}(\theta)I^{-1}(\theta)\zeta(\zeta'I^{-1}(\theta)\zeta)^{-1}]$$
$$(\text{当 }\theta\in\Theta). \tag{4.9}$$

证 沿用定理 4.1 的记号和结果，有

$$E_\theta\{(\hat{\theta}(X)-\theta)Z'(X,\ \theta)\}=\mathrm{cov}_\theta(\hat{\theta},\ Z)\triangleq D(\theta)$$
$$=I_p+\widetilde{D}(\theta).\qquad(4.10)$$

由 Schwarz 不等式可得

$$[\zeta'(I_p+\widetilde{D}(\theta))I^{-1}(\theta)\zeta]^2=E_\theta[\zeta'(\hat{\theta}-\theta)Z'(x,\ \theta)I^{-1}(\theta)\zeta]$$
$$\leqslant\mathrm{Var}_\theta(\zeta'(\hat{\theta}-\theta))\cdot\mathrm{Var}_\theta(\zeta'I^{-1}(\theta)Z(x,\ \theta))$$
$$=\zeta'\mathrm{VAR}_\theta(\hat{\theta})\zeta\cdot\zeta'I^{-1}(\theta)\zeta.\qquad(4.11)$$

从而(4.9)式成立.

二、C-R 不等式等号成立的充要条件

定理 4.3 假设定理 4.1 的条件成立，并进一步假设 $k=p$，对每个 x，有 $\frac{\partial}{\partial\theta}f(x,\ \theta)$ 关于 θ 连续，$\frac{\partial G}{\partial\theta}=D(\theta)$ 非奇、连续($\theta\in\Theta$)．且对每个 $\theta_0\in\Theta$，存在 θ_0 的邻域 $V_{\theta_0}\subset\Theta$ 使 $I(\theta)$ 在 V_{θ_0} 上有上界，则 $G(\theta)$ 的无偏估计方差阵处处达到下界((4.1)式成立)的充要条件是存在 μ 零测集 N 及 $\psi(\theta)$、$r(x)$，使

$$f(x,\ \theta)=\beta(\theta)r(x)e^{\psi^\tau(\theta)\hat{G}(x)}\quad(\text{当 }x\overline{\in}N,\ \theta\in\Theta).\qquad(4.12)$$

这时必有 $\beta(\theta)>0$ 有连续一阶偏导数，$\left(\frac{\partial\psi'(\theta)}{\partial\theta}\right)$ 非奇，关于 θ 连续，$r(x)>0$.

证 根据定理 4.1 的第二部分，(4.1)式成立的充要条件为：存在 $B(\theta)$(满秩)使一切 $\theta\in\Theta$ 有

$$\hat{G}-G(\theta)=B(\theta)Z(x,\ \theta),\quad a.s.P_\theta.\qquad(4.13)$$

且

$$B(\theta)=D(\theta)I^{-1}(\theta).\qquad(4.14)$$

根据(4.4)，当 $\theta\in\Theta$ 时，皆有

$$(\hat{G}-G(\theta))Z'(x,\ \theta)=B(\theta)Z(x,\ \theta)Z'(x,\ \theta),\ \ a.s.P_\theta.\qquad(4.15)$$

由于 $f(x,\ \theta)>0$ (当 $x\in\mathcal{X}$、$\theta\in\Theta$)．故 $P_\theta\equiv P_{\theta_0}$．根据(4.15)式，对任意确定的 $\theta_0\in\Theta$ 及 $V_{\theta_0}\subset\Theta$，当 $\theta\in V_{\theta_0}$ 时，$I(\theta)$ 及 $D^{-1}(\theta)$ 有

$$E_{\theta_0}(\hat{G}(X)-G(\theta))Z'(x,\ \theta)=B(\theta)E_{\theta_0}Z(x,\ \theta)Z'(x,\ \theta),$$

再由于 $I(\theta)>0(\theta\in\Theta)$ 及共同支撑条件知：$Z(x,\ \theta)$ 是非退化的，故 $E_{\theta_0}Z(x,\ \theta)Z'(x,\ \theta)>0$，因而

$$B(\theta)=E_{\theta_0}\{(\hat{G}-G(\theta))Z'(X,\ \theta)\}\{E_{\theta_0}Z(X,\ \theta)Z'(X,\ \theta)\}^{-1}. \tag{4.16}$$

由于

$$\begin{aligned}&|Z(x,\ \theta)Z'(x,\ \theta)|\\&=|B^{-1}(\theta)(\hat{G}(x)-G(\theta))(\hat{G}(x)-G(\theta))'B^{-1}(\theta)|\\&\leqslant k_1|B^{-1}(\theta)|^2(|\hat{G}(x)|^2+|G(\theta)|^2);\end{aligned} \tag{4.17}$$

$$\begin{aligned}&|(\hat{G}-G(\theta))Z'(x,\ \theta)|\\&=|(\hat{G}-G(\theta))(\hat{G}-G(\theta))'(B^{-1}(\theta))'|\\&\leqslant k_2|B^{-1}(\theta)|(|\hat{G}|^2+|G(\theta)|^2).\end{aligned} \tag{4.18}$$

其中 $|A|$ 表示 A 阵各元素的绝对值最大者，k_1、k_2 是常数．由于 $\mathrm{VAR}_{\theta_0}(\hat{G})$ 存在有限，故根据(4.14)及 $I(\theta)$ 的假设知：$B(\theta)$ 在 V_{θ_0} 上有界，由(4.17)～(4.18)、控制收敛定理及 $Z(x,\ \theta)$ 与 $f(x,\ \theta)$ 关于 θ 连续，就得到

$$\begin{aligned}&\lim_{\theta\to\theta_0}B(\theta)\\&=E_{\theta_0}[(\hat{G}-G(\theta_0))Z'(x,\ \theta_0)]\{E_{\theta_0}Z(x,\ \theta_0)Z(x,\ \theta_0)\}^{-1}\\&=D(\theta_0)I^{-1}(\theta_0).\end{aligned} \tag{4.19}$$

此结果说明了 $B(\theta)$ 关于 θ 连续．

充分性：若(4.12)式成立，在 N^c 中任取两点 $x_1\neq x_2$，因

$$f(x,\ \theta)>0,$$

故 $\beta(\theta)>0$（当 $\theta\in\Theta$）．同时

$$\begin{aligned}&\log f(x_1,\ \theta)-\log f(x_2,\ \theta)\\&=\log r(x_1)-\log r(x_2)+\psi^{\tau}(\theta)(\hat{G}(x_1)-\hat{G}(x_2)).\end{aligned} \tag{4.20}$$

由于 $Z(x,\ \theta)$ 存在，对每个 x 关于 θ 连续．故 $\frac{\partial\psi}{\partial\theta}$ 存在关于 θ 连续．从而 $\beta(\theta)$ 有一阶连续偏导数．此时

$$Z(x,\ \theta)=\frac{\partial}{\partial\theta}\log\beta(\theta)+\frac{\partial\psi^{\tau}(\theta)}{\partial\theta}\hat{G}(x); \tag{4.21}$$

$$0<I(\theta)=\left(\frac{\partial\psi(\theta)}{\partial\theta}\right)'VAR_\theta\hat{G}(X)\frac{\partial\psi(\theta)}{\partial\theta}. \tag{4.22}$$

故 $\frac{\partial\psi(\theta)}{\partial\theta}$ 满秩. 再由定理 4.1 及(4.21)式知 C-R 不等式(4.1)在 Θ 上处处有等号成立.

必要性：根据定理 4.1，若(4.1)在 Θ 上处处有等号成立，则(4.13)对 $\theta\in\Theta$ 皆成立，从而 $B(\theta)$ 非奇、连续. 记

$$M=\{(x,\ \theta):\hat{G}(x)-G(\theta)\neq B(\theta)Z(x,\ \theta)\}. \tag{4.23}$$

以 L 表 Θ 上的 Lebesgue 测度，由(4.13)及 $\mathscr{P}\equiv\mu$ 知下式成立：

$$\int_\Theta\int_{\mathscr{X}} I_M(x,\ \theta)\,d\mu\,dL=0.$$

由 Fibini 定理知：存在 μ 零测集 N，使当 $x\overline{\in}N$ 时，$L(M^x)=0$（M^x 表 M 在 x 处的切口）. 又由于 $B(\theta)$、$G(\theta)$、$Z(x,\ \theta)$ 关于 θ 连续，故 $M^x=\phi$，所以当 $x\overline{\in}N$ 时，(4.13)对所有的 $\theta\in\Theta$ 成立. 因此，通过解微分方程(4.13)，即得

$$\log f(x,\ \theta)=\psi^\tau(\theta)\hat{G}(x)+W(\theta)+h(x)$$
$$(\text{当 } x\overline{\in}N,\ \theta\in\Theta), \tag{4.24}$$

其中

$$\frac{\partial\psi^\tau(\theta)}{\partial\theta}=B^{-1}(\theta),\quad \frac{\partial W(\theta)}{\partial\theta}=-B^{-1}(\theta)G(\theta).$$

对(4.24)的两边取反对数，即得

$$f(x,\ \theta)=\beta(\theta)e^{\psi^\tau(\theta)\hat{G}(x)}r(x),$$

其中 $\beta(\theta)=e^{W(\theta)}>0$，$r(x)=e^{h(x)}$，$\psi(\theta)$ 皆满足要求. 定理证毕.

定理 4.4 若分布族 $\mathscr{P}$ 及可估函数 $G(\theta)=\theta$ 及其无偏估计 $\hat{G}(x)$ 皆满足定理 4.1 的条件，那么

i) 若 $I(\theta)=V>0$（与 θ 无关），则 $\mathrm{VAR}\hat{G}(x)$ 处处达到 C-R 下界的充要条件为 $\hat{G}(X)\sim N(\theta,\ V^{-1})$.

ii) 若 $I(\theta)=\left(n\theta_i^{-1}\delta_{ij}+n\left(1-\sum_{i=1}^{p-1}\theta_i\right)^{-1}\right)_{p\times p}$，其中 $\theta_i>0$ ($i=1,\ 2,\ \cdots,\ p$)，$\sum_{i=1}^{p}\theta_i=1$. 则 $\mathrm{VAR}_\theta\hat{G}$ 处处达到 C-R 下界的充要条件为 $n\hat{G}$ 服从参数为 $\theta_1,\ \theta_2,\ \cdots,\ \theta_p$ 的多项分布.

iii) 若 $I(\theta)=(\alpha_i\theta_i^{-2}\delta_{ij})_{p\times p}$，$\alpha_i>0$，$\theta_i>0$ ($i=1,\ 2,\ \cdots,\ p$)，则 $\mathrm{VAR}_\theta\hat{G}$ 处处达到 C-R 下界为 $\hat{G}(x)$ 服从参数为 $(\alpha_1,\ \cdots,\ \alpha_p;\ \alpha_1\theta_1^{-1},\ \cdots,\ \alpha_p\theta_p^{-1})$ 的 Γ 分布.

证 充分性：根据定理 4.3，充分性是显然的．我们只证必要性：在 i) 的情况，由定理 4.3 知道：

$$f(x,\ \theta)=e^{\psi^{\tau}(\theta)\hat{G}(x)+w(\theta)+h_1(x)};\tag{4.25}$$

$$Z(x,\ \theta)=\frac{\partial\psi^{\tau}(\theta)}{\partial\theta}(\hat{G}(x)-\theta)=B^{-1}(\theta)(\hat{G}(x)-\theta).\tag{4.26}$$

因为 $D(\theta)=I$，故由(4.14)知，

$$B(\theta)=I^{-1}(\theta)=V^{-1}=\left(\frac{\partial\psi^{\tau}}{\partial\theta}\right)^{-1}.$$

故

$$\psi(\theta)=V\theta+\alpha.$$

而 $\dfrac{\partial W}{\partial\theta}=-\dfrac{\partial\psi^{\tau}}{\partial\theta}G(\theta)=-V\theta$，从而

$$W(\theta)=-\frac{1}{2}\theta'V\theta+c.$$

此时

$$\begin{aligned}f(x,\ \theta)&=\exp\left\{\theta'V\hat{G}(x)+\alpha'\hat{G}(x)-\frac{1}{2}\theta'V\theta+c+h_1(x)\right\}\\&=\exp\left\{\theta'V\hat{G}(x)-\frac{1}{2}\theta'V\theta+h(x)\right\}.\end{aligned}\tag{4.27}$$

令 $\theta'V=t'$，故

$$\frac{1}{2}\theta'V\theta=\frac{1}{2}t'V^{-1}t.$$

由指数族分布特征函数的表达式为

$$e^{\frac{1}{2}(t+iu)'v^{-1}(t+iu)-\frac{1}{2}t'v^{-1}t}=e^{iu'v^{-1}t-\frac{1}{2}u'v^{-1}u}=e^{iu'\theta-\frac{1}{2}u'v^{-1}u}.\tag{4.28}$$

利用反演公式即知 $\hat{G}(x)\sim N(\theta,\ v^{-1})$．

ii) 与 iii) 的必要性的证明与 i) 完全相似，即利用定理 4.3 求出 $\psi(\theta)$ 及 $\beta(\theta)$(或 $W(\theta)$)，再计算 $\hat{G}(x)$的特征函数，再由特征函数唯一决定分布函数。

由此定理也不难推得：若假设 $f(x,\ \theta)=h(x)e^{\theta't(x)+\psi(\theta)}>0$，则 θ 的无偏估计处处达到 C-R 下界的充要条件是 $t(X)\sim N(\theta,I)$．

三、多参数 Fisher 信息阵的性质

如同 § 2 一样，利用类似的方法可得到多参数 Fisher 信息阵 $I(\theta)=E_{\theta}(Z(x,\ \theta)Z'(x,\ \theta))$的性质如下：

(1) $I(\theta)\geqslant 0$, 且与 μ 选择无关;

(2) 如 X、Y 独立, 其分布族皆满足正则条件 i**)~iv**), 则 (X, Y) 服从的分布族也满足 i**)~iv**), 且

$$I^{(x,y)}(\theta)=I^{x}(\theta)+I^{y}(\theta).$$

(3) 假设 X 及其统计量 $T(X)$ 的分布族 $\mathscr{P}, \mathscr{P}^T$ 皆满足正则条件 i**)、ii**)、iv**) 和对任意的 $A\in\mathscr{B}_x$、$B\in\mathscr{B}_y$ 皆有

$$\int_A \frac{\partial}{\partial\theta}f(x, \theta)d\mu=\frac{\partial}{\partial\theta}\int_A f(x, \theta)d\mu,$$

$$\int_B \frac{\partial}{\partial\theta} f^T(t, \theta)d\mu^T=\frac{\partial}{\partial\theta}\int_B f^T(t, \theta)d\mu^T,$$

则有 $I^{T(x)}(\theta)\leqslant I^{x}(\theta)$, 如果

$$Z(x, \theta)=\frac{\partial}{\partial\theta}\log f(x, \theta), \quad \widetilde{Z}(t, \theta)=\frac{\partial}{\partial\theta}\log f^T(t, \theta),$$

则 $I^T(\theta)=I^x(\theta)$ 的充要条件是 T 为 θ 的充分统计量, 此处 $f^T(t,\theta)$ 表示 $T(x)$ 的密度(关于 μ^T).

上述证明皆从略.

四、多参数的 Bh 不等式

对于多参数的情况, 同样也有 Bhattacharrya 不等式成立. 假设随机向量 $X\sim P_\theta$, $\frac{dP_\theta}{d\mu}=f(x, \theta)$, $\theta\in\Theta$, Θ 为 R 开区域. 定义

$$Z_{i_1,\cdots,i_p}(x, \theta)=(f(x, \theta))^{-1}\frac{\partial^{i_1+\cdots+i_p}}{\partial\theta_1^{i_1}\cdots\partial\theta_p^{i_p}}f(x, \theta); \quad (4.29)$$

$R=\{(i_1, \cdots, i_p), i_\nu$ 为非负整数, $1\leqslant\nu\leqslant p$, 且不全为零$\}$

在 R 上规定次序如下:

(1) 若 $\sum_{i=1}^{p} r_i<\sum_{i=1}^{p} t_i$, 则说 $(r_1, \cdots, r_p)\prec(t_1, \cdots, t_p)$;

(2) 若 $\sum_{i=\nu}^{p} r_i=\sum_{i=\nu}^{p} t_i$, $1\leqslant\nu\leqslant m$, 而 $\sum_{i=m+1}^{p} r_i<\sum_{i=m+1}^{p} t_i$, 则说

$$(r_1, \cdots, r_p)\prec(t_1, t_2, \cdots, t_p).$$

对任意正整数 N 和一组非负整数 $(I_1, I_2, \cdots, I_p)$, $\sum_{i=1}^{p} I_i>0$, 记

$$R_N(I_1, \cdots, I_p)=\Big\{(i_1, \cdots, i_p):(i_1, \cdots, i_p)\in R,\ i_\nu\leqslant I_\nu,$$

$$\nu=1, 2, \cdots, p, \sum_{\nu=1}^{p} i_\nu\leqslant N\Big\}.$$

$$Z(x, \theta)\triangleq Z_{N, I_1, \cdots, I_p}(x, \theta)=\{Z_{i_1, \cdots, i_p}(x, \theta)(i_1, \cdots, i_p)$$
$$\in R_N(I_1, \cdots, I_p)\}, \tag{4.30}$$

其中随机向量 $Z(x, \theta)$ 的分量按前面所规定的 $(i_1, \cdots, i_p)$ 的次序排列.

设 $G(\theta)=(g_1(\theta), \cdots, g_k(\theta))'$ $(k\leqslant p)$，记

$$g_\nu^{(f)}(\theta)=\frac{\partial^{i_1+\cdots+i_p}}{\partial\theta_1^{i_1}\partial\theta_2^{i_2}\cdots\partial\theta_p^{i_p}}g_\nu(\theta),$$

此处 f 为 $(i_1, \cdots, i_p)$ 在 $R_N(I_1, \cdots, I_p)$ 中的次序. 记

$$\tilde{g}_\nu=(g_\nu^{(1)}(\theta), \cdots, g_\nu^{(M)}(\theta))' \quad (\nu=1, \cdots, R),$$

其中 M 为 $R_N(I_1, \cdots, I_p)$ 元素的个数; 再记

$$D(\theta)=(\tilde{g}_1, \cdots, \tilde{g}_R)'_{R\times M}.$$

如果 $G(\theta)$ 可估, $\hat{G}(x)$ 是其无偏估计, 在与 § 3 类似的正则条件下, 则有

i) $$\mathrm{VAR}_\theta\hat{G}\geqslant D(\theta)V^{-1}(\theta)D^{\tau}(\theta); \tag{4.31}$$

ii) (4.31) 等号成立 $\Leftrightarrow$ 存在 $p\times N$ 满秩矩阵 $B(\theta)$，使得

$$\hat{G}(x)-G(\theta)=B(\theta)Z(x, \theta) \quad (当\ \theta\in\Theta),$$

在 (4.31) 中, $\quad V(\theta)=E_\theta(Z(x, \theta)Z'(x, \theta)).$

我们称 $D(\theta)V^{-1}(\theta)D'(\theta)$ 为 $R_N(I_1, \cdots, I_p)$ 阶 Bh 不等式下界.

例 4.1 令 $X_1, X_2, \cdots, X_n$ 为 $n(n\geqslant2)$ 个 iid. 随机变量, 遵从正态分布 $N(\mu, \sigma^2)$, 则 $\theta=(\mu, \sigma^2)'$ 的 UMVUE 估计达不到 C-R 下界, 但达到 $R_3(2, 1)$ 阶 Bh 下界. 此时,

$$f(x, \theta)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\,\sigma}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}$$

$$=(2\pi)^{-\frac{n}{2}}\sigma^{-n}e^{-\frac{n(\bar{x}-\mu)^2}{2\sigma^2}-\frac{s^2}{2\sigma^2}}.$$

此处 $\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$, $s^2=\sum_{i=1}^{n}(x_i-\bar{x})^2$.

$$Z_1(x,\ \theta)=(f(x,\ \theta))^{-1}\frac{\partial}{\partial\mu}f(x,\ \theta)=n\sigma^{-2}(\bar{x}-\mu);$$

$$\begin{aligned}Z_2(x,\ \theta)&=(f(x,\ \theta))^{-1}\frac{\partial}{\partial\sigma^2}f(x,\ \theta)\\&=-\frac{n}{2\sigma^2}+\frac{1}{2}\left(\frac{s^2}{\sigma^4}+\frac{n(\bar{x}-\mu)^2}{\sigma^4}\right);\end{aligned}$$

$$\begin{aligned}Z_3(x,\ \theta)&=(f(x,\ \theta))^{-1}\frac{\partial^2}{\partial\mu^2}f(x,\ \theta)\\&=\frac{n^2}{\sigma^4}(\bar{x}-\mu)^2-\frac{n}{\sigma^2};\end{aligned}$$

$$V(\theta)=E_\theta Z(x,\ \theta)Z'(x,\ \theta)=\begin{pmatrix}\frac{n}{\sigma^2} & 0 & 0\\ 0 & \frac{n}{2\sigma^4} & \frac{n}{\sigma^4}\\ 0 & \frac{n}{\sigma^4} & \frac{2n^2}{\sigma^4}\end{pmatrix};$$

$G(\theta)=\theta$ 的 UMVUE 为 $\hat{G}(x)=\left(\bar{x},\ \frac{s^2}{n-1}\right)'$；

$$\mathrm{VAR}_\theta(\hat{G}(X))=\begin{pmatrix}\frac{\sigma^2}{n} & 0\\ 0 & \frac{2\sigma^4}{n-1}\end{pmatrix}.$$

其 C-R 下界为

$$\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\begin{pmatrix}\frac{n}{\sigma^2} & 0\\ 0 & \frac{n}{2\sigma^4}\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}=\begin{pmatrix}\frac{\sigma^2}{n} & 0\\ 0 & \frac{2\sigma^4}{n}\end{pmatrix}.$$

显然，$\mathrm{VAR}_\theta(\hat{G}(x))$没有达到 C-R 下界．而 $R_3(2,\ 1)$阶的 Bh 不等式下界为

$$\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\end{pmatrix}V^{-1}(\theta)\begin{pmatrix}1 & 0\\ 0 & 1\\ 0 & 0\end{pmatrix}=\begin{pmatrix}\frac{\sigma^2}{n} & 0\\ 0 & \frac{2\sigma^4}{n-1}\end{pmatrix}.$$

这说明 $\mathrm{VAR}_\theta(\hat{G}(x))$达到了 $R_3(2,\ 1)$阶的 Bh 不等式下界．

问题与习题

1. 设 $r.v\ X\sim N(\mu,\sigma^2)$，如果 μ 的估计 $\hat{\mu}(x)$ 满足 $E\{(\mu-\hat{\mu}(X))^2|\mu,\sigma^2\}\leqslant\sigma^2$（当 $\mu\in R^1$），则 $\hat{\mu}(x)=x$, a、s、L. [提示：用 C-R 不等式(1.20)，解微分不等式 $b^2(\mu)+(1+b'(\mu))^2\sigma^2\leqslant\sigma^2$，得 $b(\mu)=0$. 此处

$$b(\mu)=\mu-E_\mu(\hat{\mu}(X)).$$

从而推得 $\hat{\mu}(x)=x$, a、s、L.]

2. 当 $r.v\ X$ 具有概率密度 $\frac{1}{2}e^{-|x-\theta|}(-\infty<\theta<\infty)$，试问 C-R 不等式是否成立？如果成立，试证明之.

3. 令 $x_1, x_2, \cdots, x_n$ 为 n 个 iid. 随机变量，服从如下三点分布

$$P_\theta(X=y_1)=\frac{1-\theta}{2},\ P_\theta(X=y_2)=\frac{1}{2},\ P_\theta(X=y_3)=\frac{\theta}{2}(0<\theta<1),$$

试问 C-R 不等式在这种分布下是否成立？它的无偏估计的最小方差是什么？

4. 试证明 Kiefer 不等式(1.22).

5. 设 $x_1, x_2, \cdots, x_n$ 为 n 个 iid. 随机变量，服从：a) $(0,\theta)$ 上的均匀分布；b)寿命分布 $1-e^{-(x-\theta)}(x\geqslant\theta)$. 试证：$\theta$ 的 UMVUE 皆可达到 Kiefer 不等式下界(1.23). [提示：在情况 a) 取 $d\lambda_1(h)=\frac{n+1}{\theta}\left(\frac{h}{\theta}+1\right)^n dh$ （当 $-\theta<h<0$)；在情况 b) 取 $d\lambda_1(h)=\theta e^{-\theta h}dh$. （当 $0<h<\infty$).]

6. 设 $r.v\ X$ 服从分布 $P_\theta(\theta\in\Theta)$，且有 $\frac{dP_\theta}{d\mu}=f(x,\theta)$ $(a<\theta<b)$，而且 $\frac{\partial^2 f(x,\theta)}{\partial\theta^2}$ 存在，以及

$$0=\frac{\partial^2}{\partial\theta^2}\int f(x,\theta)d\mu=\int\frac{\partial^2}{\partial\theta^2}f(x,\theta)d\mu,$$

则 Fisher 信息函数 $I(\theta)=-E_\theta\left(\frac{\partial^2}{\partial\theta^2}\log f(X,\theta)\right)$.

7. 若 $r.v\ X$ 服从极值分布，其密度函数为

$$f(x,\alpha)=\exp\{-(x-\alpha)-e^{-(x-\alpha)}\}(x\in R^1,\ \alpha\in R^1)$$

记 $\theta=e^\alpha$. 试求：i) e^{-x} 的分布函数；ii) 以 θ、α 分别作为参数的 Fisher 信息函数. 并指出这两个信息函数的关系是什么？

8. 若 $r.v\ X\sim$ 二项分布 $B(n,p)(0<p<1)$，试求 $np(1-p)$ 的 UMVUE 和此估计的方差；并指出它达到哪一阶 Bh 不等式下界？

9. 若 $r.v\ X\sim N(0,\sigma^2)$. 试求 $\frac{1}{\sqrt{2\pi}\sigma}$ 的 UMVUE 及其方差.

10. 多项分布

$$P(X_1=r_1,\ \cdots,\ X_k=r_k)=\frac{n!}{r_1!\cdots r_k!}\,p_1^{r_1}\cdots p^{r_k}\left(\sum_{i=1}^{k}r_i=n,\ \sum_{i=1}^{k}p_i=1\right)$$

及 Weibull 分布，其密度为 $\frac{r}{\beta}x^{r-1}e^{-\beta^{-1}x^r}(0<r,\ 0<\beta)$. 试求这两个分布的 Fisher 信息阵和 C-R 下界；并指出是否存在它们所含未知参数的无偏估计，而能达到 C-R 下界.

11. 试证明多参数的 Bh 不等式以及有偏估计的 Bh 不等式.

12.* 假设 $x_1,\ x_2,\ \cdots,\ x_n$ $(n\geqslant 2)$ 为 n 个 $iid.$ 随机变量，服从正态分布 $N(\mu,\ \sigma^2)$, $\mu\in R^1$, $\sigma>0$. 如果存在 σ^2 的二次型估计

$$\hat{\sigma}_n^2(x_1,\ \cdots,\ x_n)=\sum_{i<j}a_{ij}x_ix_j,$$

使得
$$E(\hat{\sigma}_n^2-\sigma^2)^2\leqslant\frac{2\sigma^4}{n+1}$$

对一切 $\sigma>0$, $\mu\in R^1$ 成立. 则

$$\hat{\sigma}_n^2=\frac{1}{n+1}\sum_{i=1}^{n}(x_i-\bar{x}_n)^2.$$

[提示：i) 利用 $\bar{x}_n=\frac{1}{n}\sum_{i=1}^{n}x_i$ 及 $s_n^2=\sum_{i=1}^{n}(x_i-\bar{x}_n)^2$ 是 $(\mu,\ \sigma^2)$ 的充分统计量. 从而满足上述不等式 σ^2 的二次型估计必具有 $a\bar{x}_n^2+bs_n^2(a\geqslant 0,\ b\geqslant 0)$ 的形式.

ii) 用多参数有偏估计的 Bh 不等式，得到类似于习题 1 的不等式，由不等式解得 $a=0$, $b=\frac{1}{n+1}$.]

13.* 设 r、v、$X\sim P_\lambda(x)$, 且 $\frac{dP_\lambda}{d\mu}=f(x,\ \lambda)\,(\lambda>0)$ 满足正则条件 i)～vi). 试证：λ 的无偏估计能达到 C-R 下界 λ 的充要条件是 $X\sim$ Poisson 分布 $\mathscr{P}(\lambda)$.

参 考 文 献

[1] H. Cramer: Mathematical Method In Statistics, Princeton (1948).

[2] C. R. Rao: Information and the accuracy attainable in the estimation of statistical parameters, Bull. Calcutta Math. Soc. 37: 81～91(1945).

[3] A Bhattacharyya: On some analogues to the amount of information an their uses in statistical estimation, Sankhya 8: 1～14(1946).

[4] A. V. Fend: On the attainment of Cramer-Rao and Bhattacharyya bounds for the variances of an estimate, Ann. Math. Statist. 30: 381～388(1959).

[5] R. A. Wijsman: On the attainment of the Cramér-Rao lower bound, Ann. statist. 1: 538～542(1973).

[6] V. M. Joshi: On the attainment of the Cramér-Rao lower bound, Ann.

Statist. 4: 998(1976).

[7] J. L. Hodge, E. L. Lehmann: Some applications of the Cramer-Rao inequallity, Second Berkely Symp. Math. Statist. Prob. I: 13～32 (1951).

[8] R. V. Hogg and A. T. Craig: *Some results on unbiased estimation*, Sankhya, Series A 24: 363～338 (1962).

[9] J. Kiefer: On minimum variance estimators, Ann. Math. Statist. 23: 627～629(1952).

[10] D. G. Chapman and H. Robbins: Minimun variance estimation without regularity assumptions, Ann. Math. Statist. 22: 581～586(1951).

[11] R. A. Fisher: Theory Of Statistical Estimation, Proc. Phill. Soc. 22: 700～715(1925).

[12] S. Kullback and R. A. Leibler: On information and sufficiency Ann. Math. Statist. 22: 79～86(1951).

[13] D. N. Shanbhag: Some charaterizations based on the Bhattacharyya matix, J. Appl. Prob. 9: 580～587(1972).

[14] B. J. N. Blight & P. V. Rao: The convergence of Bhattacharyya bounds, Biometika 61: 137～142(1974).

[15] 李国英、成平:《关于 C-R 型下界的几个结果》摘要载于《科学通报》第 2 期 75～78(1981), 全文在《数学学报》第 25 卷 76～93(1982).

[16] C. R. Blyth and D. M. Roberts: On inequalities of Cramér-Rao type and admissibility proofs, Proc. Sixth Berkely Symp. Math. Statist. Prob. I: 17～30(1972).

[17] Fabian Václar, Hannan James: On the Cramér-Rao inequality, Ann. Statist. 5: 197～205 (1977).

[18] C. N. Morris: Natural exponential families with quadratic variance functions. Ann. Statist. 10: 65～80(1982).

第四章 Bayes 估计[1)]

§1 引 言

在第一章中我们已提到过 Bayes 方法. 当时它只是作为判决函数的一种优良性准则而提出来的. 其实"Bayes 学派"(指那些鼓吹以 Bayes 方法作为全部统计推断(及判决)的合理基础的学者所组成的学派)现已成为统计学中一个很有影响的学派,它在一些根本点上与"古典统计"有分歧. 因此,有必要花费一点篇幅来作些解释.

所谓"古典统计",按照 J. O. Berger 在其最近著作[1]中表述的观点,其特点主要有二:一是它只依据样本而不考虑先验知识;另一是,古典统计推断的"可信度",一般是"事前决定的". 就是说,某项统计推断的信度如何,在进行观察以获得样本 x 以前就决定了. 无论得到的具体样本如何,都不能对此产生影响. Bayes 学派在这两个根本点上都与古典学派有原则分歧.

设统计推断的对象是参数 θ. 从古典学派看来,在得到样本 x 之前,除了 $\theta\in\Theta$(Θ 是给定的参数空间)之外,其他一无所知. θ 当然也不视为是随机变量. Bayes 学派则认为:即使在进行观测以得到样本 x 之前,人们对 θ 也会有一些知识. 这可以是由于某种理论、以往在对付同类问题时所积累的经验、或者是观察者的主观认识. 由于这种知识是在试验前得到的,故可以称为验前或先验知识(Prior information). 表述这种先验知识的最简单和方便

1) 虽然本章的标题是"Bayes 估计",但许多内容对一般判决问题也适用. 局限于 Bayes 估计问题并不带来叙述上的简化,且有一些处理上的困难. 然而,为照顾本书性质,在举例及编制习题时,我们基本上局限于估计问题.

的方法，是将 θ 视为一随机变量，而给出它(在 $\mathscr{B}_\Theta$ 上)的一个概率分布 H，叫做 θ 的先验分布(Prior distribution). 比方说，设 $A\in\mathscr{B}_\Theta$，而 $H(A)=0.9$，则表示在做试验前，我们已有 90% 的把握预言未知的 θ 落在集 A 内. 这种知识如果是确实的，当然会有助于对 θ 的推断。

从常识上看，Bayes 学派关于在考虑 θ 的推断和判决问题时必须考虑 θ 的先验知识的观点，无疑是正确的. 问题在于：在每个具体问题中这种先验知识是否存在，是否能以概率分布的形式加以确切描述，则很难给出使人信服的一般回答. 下面是一个能够作出这种回答的简单例子.

设某厂每日进行抽样检查，以估计当日生产的产品的废品率 θ. 就特定的一日而言，θ 当然只是一个确定的未知数，无随机性可言. 但逐日的废品率会有些波动(在此，我们把这种波动的原因假设为纯随机性的，总的生产条件保持不变)，故从较长一段时期看，就可以把 θ 视为一随机变量. "特定一日的废品率"则是 θ 的一个"实现". 如果我们在一个较长时期对逐日的废品率(或其估计值)作了记录，则所积累的资料也可以大体上定出 θ 的先验分布，在推断 θ 时考虑到这种先验知识，显然是有益的. 比方说，若 θ 的先验分布的概率绝大部分集中在 $\theta=0$ 附近，则表示从历史上看该厂的废品率一向很低. 这时，若我们要通过从某一日生产的产品中去抽样以估计废品率 θ，则当样品中的废品较多时，θ 的先验知识将使我们把 θ 的估计值(仅仅基于当日样品的估计值)往低的方向适当调整一些，而这显然是合理的.

另一方面，若被估计的 θ 是某个物理常数(如光的速度)，或是一个大铁矿的含铁百分率，则类似于上面那种解释就不可能，且总的说来，用 Bayes 学派的观点来处理这种问题显得勉强和不自然. 当然，Bayes 学派在谈到 θ 的先验分布 H 时，并不一定需要 θ 在通常的意义上是一个随机变量，也可以是当作试验者个人对 θ 的主观认识的一种方便的描述. Bayes 学者发展了一些方法以帮助确定这种"主观分布". 这样做是否合理和有益，自然会是一个引

起争论的问题．在 θ 可以合理地认为是一个随机变量且有一定先验知识的情况，这种知识也往往并非充足到能据以确定 θ 的先验分布的程度．因此，先验分布往往带有相当的主观性和人为性．因而非 Bayes 学派认为把这种方法作为统计推断的基础是不合理的。

Bayes 学派提出的另一条辩护理由是："不管愿意与否，每个人都自觉或不自觉地是一个 Bayes 主义者"．详细解释如下：Bayes 学者提出了一些他们认为是直观而简单合理的原则，以作为衡量和比较判决函数的优良性的根据．然后证明：只要遵守这些原则，则必存在适当的先验分布 H，使优良性准则归结为在第一章中所提到过的 Bayes 原则（确切的陈述和证明可参看[2]或[3]、[4]）．这有一定的说服力，但并不能从根本上回答下述批评，即：承认 Bayes 原则在理论上有这种普遍性，但因为没有解决在具体问题中确定先验分布这个重要问题，使它在实用上未必优于其他方法．

以上所谈的是 Bayes 学派及其批评者在第一个根本点（在考虑 θ 的推断问题时应否假定 θ 有先验分布）上的分歧．下面来讨论另一个方面，即决定一个方法的信度的"事前"和"事后"．

Neyman–Pearson 的假设检验理论与 Neyman 的区间估计理论可以算是古典方法的代表．在 Neyman–Pearson 的假设检验理论中，指定一个检验水平 α，就是犯第一种错误的概率．注意 α 是在获得样本 x 以前就指定了的．如果抽样得到 x，而零假设应被接受，则不论 x 如何，犯错误的可能性也可以称为在这个特定问题中的信度总是 $1-\alpha$．Bayes 学派则认为不然．例如，检验 $N(\theta, 1)$ 中的 $\theta\leqslant 0$，抽样一次，则得到 $x=-10^{10}$ 而接受 $\theta\leqslant 0$，其"信度"应比在得到 $x=-2$ 时接受 $\theta\leqslant 0$ 为高．用 Bayes 方法可以给这一点以自然的解释．下面的例子更生动地表明这一点．

设 $x_1, \cdots, x_n$ 为抽自 $R\left(\theta-\frac{1}{2},\ \theta+\frac{1}{2}\right)$ 的 iid. 样本，要作 θ 的置信系数为 0.95 的置信区间．取 $n=25$，依 Neyman 理论，不

难得出 $\hat{\theta}\pm0.056$ 就是这样一个置信区间，此处

$$\hat{\theta}=\frac{1}{2}\left(\min_{1\leqslant i\leqslant n}\{x_i\}+\max_{1\leqslant i\leqslant n}{}_i\{x_i\}\right).$$

这里可以把 0.95 看作是"信度"，它是在抽样前就定下了的. 现设甲、乙两人分别抽样 25 次，得

甲：$\min\{x_i\}=3.1,\ \max\{x_i\}=3.2$;

乙：$\min\{x_i\}=3.0,\ \max\{x_i\}=3.96$.

从甲得到的样本看，我们只能推知 θ 落在

$$3.2-\frac{1}{2}=2.7 \quad 及 \quad 3.1+\frac{1}{2}=3.6$$

之间. 换句话说，由于我们不巧得到这么"坏"的一组样本，我们对 θ 的了解比抽样前并不多出多少. 在这种情况下，仍认定区间 $\frac{1}{2}(3.1+3.2)\pm0.056$ 即 $(3.094,\ 3.206)$ 有"信度" 0.95 是不可信的. 反之，就乙所得的样本而言，我们有 100% 的把握断言 θ 在区间 $(3.46,\ 3.50)$ 之内. 因此，就乙而言，断言 θ 落在 $\frac{1}{2}(3.0+3.96)\pm0.056$ 即 $(3.424,\ 3.536)$ 之内的可靠程度为 95% 是荒唐可笑的. 本例突出地说明了：不考虑具体抽样结果，而在事前规定一个信度，则会导致很不合理的结果. 当然，在古典方法中时信度的解释是建立在大量重复而产生的频率的基础上. 但在"一次性"的场合，这种解释是不适用的.

相反，在 Bayes 方法之下，推断的信度有一种"事后"的性质，因此较为合理和自然. 这个差别的产生，是因为 Bayes 学者对于样本的作用及推断的意义的理解与古典方法根本不同.

Bayes 学派认为，关于参数 θ 的知识，完全包含于其概率分布中. 是抽样前，知道 θ 的先验分布为 $H(\theta)$，这包含了到当时为止，我们对 θ 的全部认识. 得出样本 x 后，θ 有条件分布 $H(\theta|x)$，这概括了我们对 θ 的新认识. 因此，对 Bayes 学者而言，获得样本 x 的净效果只在于将 θ 的分布由 $H(\theta)$ 转换为 $H(\theta|x)$. 由于后者是在获得 x 以后定出的，故通称为"后验分布"(Posterior distri-

bution). 对 θ 所作的任何推断及其信度,都完全基于且仅基于这个后验分布 $H(\theta|x)$. 这样, 精度将与样本 x 有关(当然, 也与先验分布 H 有关),因而必然是"事后"的. 拿前面所举的 $R\left(\theta-\frac{1}{2}, \theta+\frac{1}{2}\right)$的区间估计的例子来说, 如果假定事前对 θ 所知甚少因而可以比较合理地认定它在某一充分大的区间$(-A, A)$内"同等可能",即认为 θ 的先验分布为 $R(-A, A)$, A 充分大,则不难算出,在得到甲、乙的样本后,各自算出的后验分布为:

甲: $R(2.7, 3.6)$; 乙: $R(3.46, 3.50)$. 因此对于由古典方法算出的区间,甲、乙用 Bayes 法分别算出其可信度为:

甲: $(3.206-3.094)/(3.6-2.7)=0.124$; 乙: 1. 这当然是一个更易于为人们接受的合理解释. 但是也要说明: 虽然一般说来, 信度的事后确定较事前确定为合理因而在这方面对古典方法的批评是中肯的. 但迄今并没有找到一种在各方面看来完全合理和有效的确定事后信度的方法,包括 Bayes 方法在内. 问题在于: 据以确定事后信度的后验分布 $H(\theta|x)$依赖于先验分布 $H(\theta)$,而后者如前所述, 无论从其基础与确定方法上说, 都存在很大问题.

在这里还要提到: 在第一章中,我们曾把 Bayes 方法理解为判决函数的一种优良性准则. 这里我们又把 Bayes 方法解释为由 $H(\theta)$到 $H(\theta|x)$的转换以及关于 θ 的推断完全基于 $H(\theta|x)$这个原则. 在下节中要证明这两个解释是一致的.

以上是关于 Bayes 方法的一个简单的评述, 更细致的论述可参看有关专著,如[1]、[5]及[6]. 不论如何,有一点可能是统计学界全都能同意的, 即 Bayes 学派确实已成长为统计中一个很有影响和很有力量的学派, 其势头看来仍在增长. 非 Bayes 学派即使对 Bayes 学派的基本出发点持保留以至批评的态度, 但他们也使用 Bayes 方法去处理一些统计理论问题,因为事实证明, Bayes 方法已成为理论统计的一个重要工具, 它可以作为理论论证及寻找某种最优解的有力手段.

§2 基本定义与后验风险最小原理

本节的目的是引进 Bayes 方法的基本概念. 在这方面, 局限于点估计并不能带来多少方便, 因此我们就一般的统计判决问题来讨论.

一、若 干 定 义

设变量 X 的样本空间为$(\mathscr{X}, \mathscr{B}_x)$, 分布族为 $\mathscr{P}=(P_\theta, \theta\in\Theta)$, $\mathscr{B}_\Theta$ 为 Θ 的某些子集构成的 σ-域, 行动空间为$(A, \mathscr{B}_A)$, 损失函数为 $L(\theta, a)$, 定义于 $\Theta\times A$. 这是一个非负的、$\mathscr{B}_\Theta\times\mathscr{B}_A$-可测的函数. 对于任一判决函数 δ, 以 $R(\theta, \delta)$ 记其风险函数.

定义 2.1 $\mathscr{B}_\Theta$ 上任一概率测度 H 称为是 θ 的一个先验分布; 又若 δ 为一判决函数, 则

$$R_H(\delta)=\int_\Theta R(\theta, \delta)dH(\theta) \tag{2.1}$$

称为 δ 在先验分布 H 之下的 Bayes 风险.

以 $\mathscr{D}$ 记一族判决函数. 例如 $\mathscr{D}$ 可以是一切判决函数或一切非随机化判决函数, 等等.

定义 2.2 若存在 $\delta_H\in\mathscr{D}$, 使

$$R_H(\delta_H)=\inf\{R_H(\delta):\delta\in\mathscr{D}\}$$

则称 δ_H 为先验分布 H 之下的 Bayes 解. 在点估计问题中, Bayes 解常称为 Bayes 估计.

在 Bayes 问题中, θ 有某种概率分布 H. 而当给定 θ 时, 我们又知道了 X 的条件分布 P_θ. 这就足以确定 (θ, X) 的联合分布 P^*, 它是 $\mathscr{B}_\Theta\times\mathscr{B}_\mathscr{X}$ 上的一个概率分布, 由公式

$$P^*(C)=\int_\Theta P_\theta(C_\theta)dH(\theta), \quad C\in\mathscr{B}_\Theta\times\mathscr{B}_\mathscr{X} \tag{2.2}$$

定义. 此处 $C_\theta=\{x:x\in\mathscr{X};\ (\theta, x)\in C\}$. 在这个联合分布下, X 的边缘分布 P 由下式给出:

$$P(A)=P(X\in A)=P^*(\Theta\times A)=\int_\Theta P_\theta(A)dH(\theta),\quad (2.3)$$

其中 $A\in\mathscr{B}_x$.

现在我们可以给出极其重要的"后验分布"的定义. 既然 (θ, X) 作为随机变量有一定的分布 P^*, 我们就可以考虑在给定 X 下 θ 的条件分布. 这个条件分布就称为 θ 的后验分布.

定义 2.3 设给定 X 时 θ 的(正则)条件概率分布存在, 就是说, 若存在一个定义于 $\mathscr{B}_\Theta\times\mathscr{X}$ 上的函数 $H(\cdot|\cdot)$, 满足以下两条件:

(1) 对固定的 $x\in\mathscr{X}$, $H(\cdot|x)$ 为 $\mathscr{B}_\Theta$ 上的一概率测度. 对固定的 $B\in\mathscr{B}_\Theta$, $H(B|\cdot)$ 为 $\mathscr{B}_x$-可测函数.

(2) 对任何 $B\in\mathscr{B}_\Theta$, $C\in\mathscr{B}_x$, 有

$$P^*(B\times C)=\int_C H(B|x)dP(x),$$

(P^*、P 分别由(2.2)、(2.3)给出)则称 $H(\cdot|x)$ 为在得到样本 x 的条件下 θ 的后验分布(注意后验分布也与 θ 的先验分布 H 有关).

在 $(\Theta, \mathscr{B}_\Theta)$ 为欧氏这个最常见的情况下, 根据第一章定理 5.1, 上述定义中的 $H(\cdot|\cdot)$ 存在. 在一个特定的也是最重要的情况下, 上述后验分布可明确写出. 设 $\mathscr{P}=(P_\theta, \theta\in\Theta)\ll\mu$, μ 为 $\mathscr{B}_x$ 上的 σ-有限测度, 记 $f(x, \theta)=dP_\theta(x)/d\mu$. 则不难直接验证[1)]

$$H(B|x)=\int_B f(x, \theta)dH(\theta)\Big/\int_\Theta f(x, \theta)dH(\theta),\ B\in\mathscr{B}_\Theta \quad (2.4)$$

$\left[\text{当}\int_\Theta f(x, \theta)dH(\theta)=0\text{ 或 }\infty\text{ 时, 取 }\mathscr{B}_\Theta\text{ 确定的概率分布, 如 }H(\cdot)\text{, 作为 }H(\cdot|x)\right]$ 易见若 ν 为 $\mathscr{B}_\Theta$ 上的 σ-有限测度, $H\ll\nu$, 且 $dH(\theta)/d\nu=h(\theta)$, 则对任何 x 有 $H(\cdot|x)\ll\nu$, 且

$$dH(\theta|x)/d\nu=f(x, \theta)h(\theta)\Big/\int_\Theta f(x, \varphi)h(\varphi)d\varphi.\quad (2.5)$$

---

1) 在此有一个细节需要提及, 在验证中需要明确: $f(x, \theta)$ 可取为 $\mathscr{B}_x\times\mathscr{B}_\Theta$-可测函数. 这一点不难从 Radom-Nikodym 定理的证明中看出(还要用到条件: 对任何 $C\in\mathscr{B}_x$, $P_\theta(C)$ 为 θ 的 $\mathscr{B}_\Theta$-可测函数).

注意，虽然 Bayes 解的定义 2.2涉及一定的损失函数，但导致后验分布这个重要概念的过程，则与行动空间、损失函数完全无关．在 4.1 中我们曾着重指出：在 Bayes 学派看来，获得样本 x 的效果只在于将我们关于 θ 的认识由 $H(\cdot)$ 转换为 $H(\cdot|x)$．以后所作的任何推断，只基于 $H(\cdot|x)$．下面看一个简单例子：

例 2.1 为估计某事件的概率 θ，作 n 次独立试验，发现该事件出现 X 次．关于 θ 的先验分布，采用“同等无知”(equal ignorance)的原则，即先天地认为 θ 取[0, 1]内任何值有同等可能，就是说，取 θ 的先验分布 H 为 $R(0, 1)$．因为在给定 θ 时 X 的分布为二项分布 $B(n, \theta)$，按(2.5)式，不难算出给定 $X=x$ 时，θ 的后验密度(对 L 测度)为

$$h(\theta|x)=(n+1)\binom{n}{x}\theta^x(1-\theta)^{n-x}$$

(当 $\theta\in[0, 1]$，他处为 0)．这是一个峰点在 x/n 处的单峰分布，在做试验前，我们只知道 θ 在[0, 1]各处有同样的可能性，经过试验得到 x 后，θ 仍有可能取[0, 1]内的任何值，但已不是同等可能，而是较集中在点 x/n 附近．这样，总的说来，关于 θ 的了解，比试验前要确切一些

至于怎样将 $h(\theta|x)$ 用于 θ 的统计推断问题，那当然依问题的性质和使用者的选择而定．例如：

(1) 若要作 θ 的点估计，可以使用 $h(\theta|x)$ 的最大值点即 x/n．这就是通常使用的估计量，也可以使用后验分布的均值，为 $(x+1)/(n+2)$．

(2) 若要作 θ 的区间估计，则可以选择一个尽可能短的区间，但包含 $1-\alpha$ 这么多的后验概率．由单峰性，不难知道这个区间 $[a_x, b_x]$ 是通过以下方程定出的：

$$h(a_x|x)=h(b_x|x),\ \int_{a_x}^{b_x}h(\theta|x)d\theta=1-\alpha.$$

这里的 $1-\alpha$ 并不就是 Neyman 意义下的置信系数，但这种“事后精度”式的作法给人以更为合理的印象．

(3) 若要检验假设 $\theta\leqslant\theta_0$, 我们可以算出它的后验概率

$$\int_0^{\theta_0} h(\theta|x)d\theta.$$

可以在事先选定一个界限 α, 当这个概率小于 α 时,就否定假设. 在这里,α 与 Neyman-Pearson 理论中的检验水平也没有联系.

这个简单例子表明 Bayes 方法与古典方法确有根本的不同.

二、后验风险最小的原则

沿用前面的记号, 并设后验分布 $H(\cdot|\cdot)$存在. 设样本为 x 而采取行动 a, 遭受的损失为 $L(\theta, a)$. 因为在得到 x 的条件下, θ 的条件分布为 $H(\cdot|x)$, 可知在得到 x 的条件下, 因采取行动 a 而遭受的条件平均损失为 $\int_\Theta L(\theta, a)H(d\theta|x)$. 它称为后验风险(Posterior risk).

定义 2.4 设 δ 为非随机化判决函数,则称

$$R_H(\delta, x)=\int_\Theta L(\theta, \delta(x))H(d\theta|x) \tag{2.6}$$

为得到样本 x 时, δ 的后验风险. 当然,后验风险还与损失函数及先验分布有关.

在随机化判决函数 $\delta(\cdot|\cdot)$的场合,后验风险相应地定义为

$$R_H(\delta, x)=\int_\Theta\left[\int_A L(\theta, a)\delta(da|x)\right]H(d\theta|x). \tag{2.7}$$

易见

$$\begin{aligned}R_H(\delta)&=E^*[L(\theta, \delta(X))]=E^*[E^*(L(\theta, \delta(X))|X)]\\&=\int_{\mathscr{X}}\left[\int_\Theta L(\theta, \delta(x))H(d\theta|x)\right]dP(x)\\&=\int_{\mathscr{X}}R_H(\delta, x)dP(x).\end{aligned} \tag{2.8}$$

此处 E^* 表示一切计算是在(θ, X)的联合分布为 P^* 的情况下进行的. 在 δ 为随机化判决函数的场合,读者容易验证(2.8)式也成立. 现在容易证明下面的定理:

定理 2.1 设存在 $B\in\mathscr{B}_x$, $P(B)=0$, 使当 $x\bar{\in}B$ 时, 极值

$\inf\left\{\int_{\Theta} L(\theta, a)\, H\, (d\theta|x):\ a\in A\right\}$在某个 $a=a_x\in A$ 处达到. 任取 $a_0\in A$, 令

$$\delta(x)=\begin{cases} a_x, \text{当 } x\overline{\in} B; \\ a_0, \text{当 } x\in B, \end{cases} \tag{2.9}$$

则当 δ 为 $(\mathscr{X}, \mathscr{B}_x)\to(A, \mathscr{B}_A)$ 可测时, 它就是在一切判决函数类(包括随机化的)中的 Bayes 解.

证 设 $\delta^*(\cdot|\cdot)$为任一判决函数, 则当 $x\overline{\in} B$ 时, 有

$$\begin{aligned} R_H(\delta^*, x) &= \int_{\Theta}\left[\int_A L(\theta, a)\delta^*(da|x)\right]H(d\theta|x) \\ &= \int_A\left[\int_{\Theta} L(\theta, a)H(d\theta|x)\right]\delta^*(da|x) \\ &\geqslant \int_A R_H(\delta, x)\delta^*(da|x) = R_H(\delta, x). \end{aligned} \tag{2.10}$$

因为 $P(B)=0$, 由(2.10)、(2.8)可立得 $R_H(\delta^*)\geqslant R_H(\delta)$. 由 δ^* 的任意性, 即证明了 δ 为 Bayes 解.

这个定理显示一个重要事实, 即在求 Bayes 解时, 可以只考虑非随机化的判决函数. 当然, 这个结论是建筑在前述极值能(几乎处处)达到的基础上. 然而, 即使这一点不对, 但局限于非随机化的判决函数仍然正确. 事实上, 任给随机化的判决函数 δ^*, 将样本空间 $\mathscr{X}$ 分为两部分 $\mathscr{X}_1$ 与 $\mathscr{X}_2$ 如下: 当 $x\in\mathscr{X}_1$ 时, 极值

$$\inf\left\{\int_{\Theta} L(\theta, a)H(d\theta|x):\ a\in A\right\}$$

能在某点(a_x)处达到, 当 $x\in\mathscr{X}_2$ 时, 极值不能达到. 这时, 当 $x\in\mathscr{X}_1$ 时, 有

$$\begin{aligned} R_H(\delta^*, x) &= \int_A\left[\int_{\Theta} L(\theta, a)H(d\theta|x)\right]\delta^*(da|x) \\ &\geqslant \int_A\left[\int_{\Theta} L(\theta, a_x)H(d\theta|x)\right]\delta^*(da|x) \\ &= \int_{\Theta} L(\theta, a_x)H(d\theta|x). \end{aligned} \tag{2.11}$$

当 $x\in\mathscr{X}_2$ 时, 对任何 $a\in A$, 有

$$\int_\Theta L(\theta,\ a)H(d\theta|x)$$
$$>\inf\left\{\int_\Theta L(\theta,\ a')H(d\theta|x):\ a'\in A\right\}\triangleq q(x),$$

这将有 $$R_H(\delta^*,\ x)>q(x)\quad (x\in\mathscr{X}_2).$$

故对任何 $x\in\mathscr{X}_2$，可找到 $a_x\in A$，导致

$$R_H(\delta^*,\ x)>\int_\Theta L(\theta,\ a_x)H(d\theta|x) \tag{2.12}$$

定义非随机化的判决函数 $\delta(x)=a_x$，则根据(2.11)与(2.12)，有

$R_H(\delta,\ x)\leqslant R_H(\delta^*,\ x)$，且不等号当 $x\in\mathscr{X}_2$ 时成立．(2.13)

因而 $R_H(\delta)\leqslant R_H(\delta^*)$．这就是说，不论给出怎样的随机化判决函数 δ^*，总可以找到非随机化的判决函数 δ，其 Bayes 风险不超过 δ^* 的 Bayes 风险．

由(2.13)进一步可知，若 $P(\mathscr{X}_2)>0$，则 $R_H(\delta)<R_H(\delta^*)$．因而除非对一切 δ 都有 $R_H(\delta)=\infty$，否则，Bayes 解不可能存在．所以，在排除这个无兴趣的例外情况后，可以断言极值

$$\inf\left\{\int_\Theta L(\theta,\ a)H(d\theta|x):a\in A\right\} \tag{2.14}$$

对几乎每个 x(测度 P)都能达到，这是 Bayes 解存在的充要条件．当这个条件满足时，Bayes 解必然对几乎每个 x 都能达到这个极值．以上的讨论可总结为下面的定理，它就是著名的"后验风险最小原则"：

定理 2.2 若存在判决函数 δ，导致

$$R_H(\delta)<\infty, \tag{2.15}$$

则 Bayes 解存在的充要条件为：对几乎每个 $x\in\mathscr{X}$ (测度 P)，极值(2.14)都能达到．当这条件满足时，判决函数 δ 为 Bayes 解的充要条件是：对几乎每个 x 都有 $\delta(x)=a_x$，其中 a_x 使极值(2.14)达到．

应当指出，在上面的讨论中还遗留下来一个问题，即 a_x 作为 $(\mathscr{X},\ \mathscr{B}_{\mathscr{X}})$ 到 $(A,\ \mathscr{B}_A)$ 的可测性．要证明这一点，需要对 Θ 的拓扑结构与损失函数 L 的性质作一些假定，并牵涉一些繁琐的论证．

在此我们就不叙述这些细节了.

定理 2.2 还有这样一个解释: 就定义 4.2 的 Bayes 解而言,它可以丝毫不涉及 θ 是否能看作随机变量以及先验分布是否存在及如何确定这些问题, 这因为 Bayes 风险 $R_H(\delta)$ 可以简单地解释为风险函数 $R(\theta,\ \delta)$的一种加权平均. 但在 4.1 节中, 我们曾指出 Bayes 方法的基本点在于: 在得到样本 x 后,统计推断完全基于后验分布. 由定理 2.2 知,在条件(2.15)之下,定义 4.2 中的 Bayes 解确实只依赖于后验分布(因为后验风险只依赖于后验分布). 因此,按定义 4.2 的 Bayes 解并不违背 4.1 节中所阐述的基本原则.

三、Bayes 方法与充分性

如§1所述,Bayes 方法提出了一种新的统计推断原则. 我们希望,这种推断原则不应与一些"看来没有疑问"的原则发生冲突. 这些原则之一是充分性原则. 如在第一章§6所述, 这个原则是指当存在一个充分统计量 $t(x)$时, 统计推断可以只基于 $t(x)$. 我们来证明: Bayes 方法确实不与充分性原则冲突.

定理 2.3 沿用前面的记号, 且设 $(P_\theta,\ \theta\in\Theta)\ll\mu$, μ 为 $\mathscr{B}_x$ 上的 σ-有限测度. 则当存在一个充分统计量 t 时, 后验分布 $H(\cdot|x)$只依赖于 $t(x)$, 即有 $\widetilde{H}(\cdot|t(x))$的形状, 而 $\widetilde{H}(\cdot|t)$就是在给定 t 时 θ 的条件分布.

证 记 $f(x,\ \theta)=dP_\theta(x)/d\mu$, 根据因子判别定理,有

$$f(x,\ \theta)=g(t,\ \theta)h(x). \tag{2.16}$$

本来(2.16)只是当 θ 固定时,对 x 几乎处处(测度 P_θ)成立. 但在用公式(2.4)计算后验分布时, 对 $f(x,\ \theta)$的取法并无限制, 故可以就取(2.16)的右边为 $f(x,\ \theta)$以用于(2.4). 不失普遍性我们可设对一切 $x\in\mathscr{X}$ 有 $0<h(x)<\infty$. 记

$$C=\left\{t:\int_\Theta g(t,\ \theta)dH(\theta)=0\text{ 或 }\infty\right\},$$

按公式(2.4),有

$$
H(B|x)=\begin{cases}\int_B g(t,\ \theta)dH(\theta)\Big/\int_\Theta g(t,\ \theta)dH(\theta), \\ \qquad\qquad \text{当 } t(x)\overline{\in} C; \\ H(B), \quad \text{当 } t(x)\in C.\end{cases}
$$

它显然只依赖于 $t(x)$. 这证明了定理的前一结论. 后一结论直接由第一章(5.18)式可得出.

这个定理的含义是清楚的: 如果 t 是一个充分统计量, 转移到 t 的概率空间$(\mathscr{T},\ \mathscr{B}_{\mathscr{T}},\ \mathscr{P}^t)$. 有了样本 x_0 后, 算出 $t(x_0)=t_0$, 而丢掉 x_0, 则在 $t=t_0$ 的条件下算出的 θ 的条件(即后验)分布就是在给定 $X=x_0$ 的条件下所算出的 θ 的条件分布. 因此基于 x_0 或基于 t_0 所能作的 Bayes 推断是一样的.

可以证明, 在一定的条件下定理 2.3 的逆也成立, 这可参看[7]. 不过, 应当指出, 在[7]中对这个结果的陈述和证明, 从数学的严格性来看, 却存在着显然的缺陷, 这种缺陷可以弥补, 但牵涉一些复杂的细节, 此处从略.

§3 平方及绝对值损失下的 Bayes 估计

后验风险最小原则将 Bayes 解归结为寻求 (2.14) 的极值点. 具体求解当然与分布族 $\mathscr{P}$、先验分布 H、损失函数 L 及所得样本有关. 特别是 L 的形式关系很大, 当 L 有比较简单的形状, 尤其是 L 为平方损失这个重要情况时, 可以得出具体的公式.

一、平方损失下的 Bayes 估计

设 g 为定义于参数空间 Θ 上的实值函数, 行动空间为$(-\infty,\ \infty)$, 损失函数为

$$
L(\theta,\ a)=\lambda(\theta)[g(\theta)-a]^2 \quad (0<\lambda(\theta)<\infty). \tag{3.1}
$$

损失函数的这种形状表示问题的性质是找 $g(\theta)$ 的点估计.

引理 3.1(Blackwell, Girshick[8]) 设 L 给定如 (3.1), $\boldsymbol{\nu}$ 为 $\mathscr{B}_\Theta$ 上的一个概率测度, 而

$$f(a)=\int_{\Theta}L(\theta,\ a)d\nu(\theta),\tag{3.2}$$

且存在 $a_1\neq a_2$ 导致 $f(a_1)<\infty>f(a_2)$，则对一切 a，有 $f(a)<\infty$，且

$$\int_{\Theta}\lambda(\theta)d\nu(\theta)<\infty.\tag{3.3}$$

证 为确定计，设 $a_1<a_2$. 先设 $a_1<a<a_2$. 则存在 $p\in(0,1)$，使 $a=pa_1+qa_2$，$q=1-p$. 易见

$$(g(\theta)-a)^2\leqslant p(g(\theta)-a_1)^2+q(g(\theta)-a_2)^2.$$

因而由(3.2)得

$$f(a)\leqslant pf(a_1)+qf(a_2)<\infty.$$

现于 $(a_1,\ a_2)$ 内找四个点 a_3、a_4、a_3'、a_4'，使

$$a_3\neq a_4,\quad a_3'\neq a_4',\quad 0\neq a_3+a_4\neq a_3'+a_4'\neq 0,$$

则由关系式

$$(a_3-a_4)[(a_3+a_4)-2g(\theta)]=(a_3-g(\theta))^2-(a_4-g(\theta))^2$$

及 $f(a_3)<\infty$，$f(a_4)<\infty$，$a_3\neq a_4$，知

$$-\infty<\int_{\Theta}\lambda(\theta)[(a_3+a_4)-2g(\theta)]d\nu(\theta)<\infty.\tag{3.4}$$

同理得

$$-\infty<\int_{\Theta}\lambda(\theta)[(a_3'+a_4')-2g(\theta)]d\nu(\theta)<\infty.\tag{3.5}$$

由(3.4)、(3.5)、$a_3+a_4\neq a_3'+a_4'$ 及 $\lambda(\theta)>0$，得(3.3). 由(3.3)、(3.4)得

$$-\infty<\int_{\Theta}\lambda(\theta)g(\theta)d\nu(\theta)<\infty.$$

再由 $f(a_1)<\infty$，即得

$$\int_{\Theta}\lambda(\theta)g^2(\theta)d\nu(\theta)<\infty,$$

因而证明了

$$-\infty<\int_{\Theta}\lambda(\theta)g^i(\theta)d\nu(\theta)<\infty\quad(i=0,\ 1,\ 2).\tag{3.6}$$

根据(3.6)，对任何 a，立得 $f(a)<\infty$. 引理证毕.

在对一切 a，有 $f(a)<\infty$ 成立的场合，不难得出 $f(a)$ 的唯一

极小值点为

$$a^*=\int_\Theta \lambda(\theta)g(\theta)d\nu(\theta)\Big/\int_\Theta \lambda(\theta)d\nu(\theta). \tag{3.7}$$

特别是在 $\lambda(\theta)\equiv 1$ 时,

$$a^*=\int_\Theta g(\theta)d\nu(\theta)$$

即为 $g(\theta)$ 在 θ 的分布为 ν 时的均值.

把这个引理与定理 2.1 结合,并沿用前面的记号,立得

定理 3.1 设损失函数为(3.1),先验分布为 H,并记

$$W_x(a)=\int_\Theta \lambda(\theta)(g(\theta)-a)^2 H(d\theta|x),$$

则 $g(\theta)$ 的 Bayes 估计为

$$\delta_H(x)=\begin{cases}\int_\Theta \lambda(\theta)g(\theta)H(d\theta|x)\Big/\int_\Theta \lambda(\theta)H(d\theta|x), \\ \quad \text{当对一切 } a, W_x(a)<\infty \text{ 时;} \\ a(x), \text{若 } W_x(a(x))<\infty, \text{而 } W_x(a)=\infty \\ \quad \text{当 } a\neq a(x); \\ \text{任意指定的 } a^*, \text{当对一切 } a, W_x(a)=\infty \text{ 时.}\end{cases} \tag{3.8}$$

对一切 a 有 $W_x(a)=\infty$ 的这种例子不难举出;$W_x(a)$ 只在一点不为 ∞ 的一个例子如下: $X\sim N(\theta, 1)$, $H\sim N(0, 1)$, $\lambda(\theta)=e^{\theta^2}$, $g(\theta)=e^{-\theta^2}$. 详细论证留给读者.

二、绝对值损失下的 Bayes 估计

设损失函数为

$$L(\theta, a)=\lambda(\theta)|g(\theta)-a| \quad (0<\lambda(\theta)<\infty), \tag{3.9}$$

若令 $f(a)=\int_\Theta \lambda(\theta)|g(\theta)-a|d\nu(\theta)$, 此处 ν 为 $\mathscr{B}_\Theta$ 上的一个概率测度,则不难证明与引理 3.1 完全一样的结论(这个简单证明留给读者). 再结合定理 2.1 与概率论中众所周知的下述事实: 若 Z 为一维变量而 m 为其中的位数, 则 $E(|Z-m|)\leqslant E(|Z-c|)$ 对任何 c, 即可得到下面的定理.

定理 3.2 设损失函数为(3.9),先验分布为 H. 记

$$W_x(a)=\int_\Theta \lambda(\theta)\,|g(\theta)-a|\,H(d\theta|x),$$

并在 $\int_\Theta \lambda(\theta)H(d\theta|x)<\infty$ 的情况下，在 $\mathscr{B}_\Theta$ 上定义概率测度如下：

$$\eta(B|x)=\int_B \lambda(\theta)H(d\theta|x)\Big/\int_\Theta \lambda(\theta)H(d\theta|x),\ \ B\in\mathscr{B}_\Theta, \tag{3.10}$$

则 $g(\theta)$ 的 Bayes 估计为

$$\delta_H(x)=\begin{cases} g(\theta)\text{的中位数,当}\ \theta\ \text{有分布}\ \eta(\cdot|x)\text{时,} \\ \quad \text{若对一切}\ a,\ W_x(a)<\infty; \\ a(x),\text{若}\ W_x(a(x))<\infty,\text{而当}\ a\neq a(x) \\ \quad \text{时},W_x(a)=\infty; \\ \text{任意指定的}\ a^*,\text{若对一切}\ a,\ W_x(a)=\infty. \end{cases}$$

比(3.9)更一般的损失函数

$$L(\theta,\ a)=\begin{cases}\lambda(\theta)p(g(\theta)-a),\text{当}\ a\leqslant g(\theta); \\ \lambda(\theta)q(a-g(\theta)),\text{当}\ a>g(\theta).\end{cases}$$

也可以类似地处理，此处 $p\in(0,\ 1)$ 为常数，$q=1-p$. 在此需要用到如下的引理.

引理 3.2 设 F 为一维变量 Z 的分布，$E(|Z|)<\infty$. 定义

$$G(\mu)=q\int_{-\infty}^{\mu}(\mu-x)dF(x)+p\int_{\mu}^{\infty}(x-\mu)dF(x)$$
$$(-\infty<\mu<\infty),$$

则 G 的最小值在 μ 为 Z 的 p-分位数处达到.

我们把这个引理的证明以及由之而导出的 Bayes 估计的讨论，都留给读者去完成.

上面提到的几个损失函数都是下述情况的特例：

$$L(\theta,\ a)=\lambda(\theta)W(g(\theta)-a), \tag{3.11}$$

这里 $0<\lambda(\theta)<\infty$, $W(t)$ 为 $-\infty<t<\infty$ 上的非负凸函数，且

$$\lim_{|t|\to\infty}W(t)=\infty.$$

设 ν 为 $\mathscr{B}_\Theta$ 上的一个概率分布，而令

$$f(a)=\int_{\Theta}\lambda(\theta)W(g(\theta)-a)d\nu(\theta), \tag{3.12}$$

则易证 $C=\{a:f(a)<\infty\}$ 为凸集，且若 C 不止包含一点，则 f 在 C 上为凸函数．再注意到(用 Fafou 引理) $\lim\limits_{|a|\to\infty}f(a)=\infty$ (或更一般地，若 a_0 为 C 的一个边界点且不属于 C，则 $\lim\limits_{a\to a_0}f(a)=\infty$)，可知 f 必在 C 上达到最小值．由此推出在上述凸损失下 Bayes 估计的存在性．当 W 在 $(-\infty,\infty)$ 上处处可微时，使(3.12)中的 $f(a)$ 达到极小的 a，可通过解方程

$$\int_{\Theta}\lambda(\theta)W'(g(\theta)-a)d\nu(\theta)=0 \tag{3.13}$$

而得到．

三、若干例子

例 3.1 设 $X\sim B(n,\theta)$，$0\leqslant\theta\leqslant1$，$\theta$ 的先验分布为 $R(0,1)$，损失函数为 $L(\theta,a)=(\theta-a)^2$，故问题是估计 θ．前面在例 2.1 中已指出 θ 的后验分布为 β 分布 $\beta(x+1,n-x+1)$．此分布的均值即 θ 在上述先验分布之下的 Bayes 估计为

$$\delta_H(x)=\frac{x+1}{n+2}. \tag{3.14}$$

这估计的风险函数为

$$\begin{aligned}R(\theta,\delta_H)&=E_\theta\left(\frac{x+1}{n+2}-\theta\right)^2\\&=[n\theta(1-\theta)+(1-2\theta)^2]/(n+2)^2.\end{aligned}$$

因而算出 δ_H 的 Bayes 风险为

$$R_H(\delta_H)=\int_0^1R(\theta,\delta_H)d\theta=\frac{1}{6(n+2)}.$$

通常估计 X/n 的 Bayes 风险算得为 $\frac{1}{6n}$，与通常估计比较，估计(3.14)有这样一个优点：当 $x=0$ 或 n 时，按通常估计将给出 0 或 1 为 θ 的估计值．(3.14)则给出接近于 0 或 1 的值为其估计，这往往显得更合理些．

现若取 θ 的先验分布为 β-分布 $\beta(a, b)$，即对 L 测度有密度

$$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}(\text{当 } 0<\theta<1, \text{他处为 } 0), \tag{3.15}$$

其中 $a>0$, $b>0$ 为常数. 则易见在得出样本 x 时, θ 的后验密度为 $\beta(a+x, n+b-x)$. 此分布的均值即 θ 在平方损失与先验分布 (3.15) 之下的 Bayes 估计为

$$\delta_{a,b}(x)=\frac{x+a}{n+a+b}. \tag{3.16}$$

简单计算得出 $\delta_{a,b}$ 的风险函数为

$$R(\theta, \delta_{a,b})=\sum_{i=0}^{n}\binom{n}{i}\theta^{i}(1-\theta)^{n-i}\left(\frac{i+a}{n+a+b}-\theta\right)^2$$

$$=\frac{n\theta(1-\theta)+[(a+b)\theta-a]^2}{(n+a+b)^2}. \tag{3.17}$$

$\delta_{a,b}$ 的 Bayes 风险为

$$R_H(\delta_{a,b})=\int_0^1 R(\theta, \delta_{a,b})\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}d\theta$$

$$=ab/[(n+a+b)(a+b+1)(a+b)]. \tag{3.18}$$

特别是当 $a=b=\sqrt{n}/2$ 时，得估计

$$\delta_{\sqrt{n}/2,\sqrt{n}/2}(x)=\frac{x}{n+\sqrt{n}}+\frac{\sqrt{n}}{2(n+\sqrt{n})}, \tag{3.19}$$

其风险函数为一常数

$$R(\theta, \delta_{\sqrt{n}/2,\sqrt{n}/2})=\frac{n}{4(n+\sqrt{n})^2}. \tag{3.20}$$

本例中的先验分布——β 分布有这样一个特点，就是不论样本 x 如何，后验分布总是与先验分布同属于一个族，在此为 β 分布族. 这种先验分布族叫共轭先验分布族 (family of conjugate prior distributions). 它在理论上有重要意义. 我们将在下节对这种先验分布族作更仔细的讨论.

例 3.2 设 $X=(X_1, \cdots, X_n)$, $X_1, \cdots, X_n$ 为抽自 $N(\theta, 1)$ 的 iid. 样本，损失函数为 $L(\theta, a)=(\theta-a)^2$，先验分布为 $N(0, \tau^2)$，要求 θ 的 Bayes 估计.

记 $\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$，写出$(\theta, X_1, \cdots, X_n)$的联合密度

$$\frac{1}{\sqrt{2\pi}\,\tau}\exp(-\theta^2/2\tau^2)\left(\frac{1}{\sqrt{2\pi}}\right)^n\exp\left(-\sum_{i=1}^{n}(x_i-\theta)^2/2\right),$$

并注意

$$\theta^2/\tau^2+\sum_{i=1}^{n}(x_i-\theta)^2=\frac{1+n\tau^2}{\tau^2}\left(\theta-\frac{n\tau^2\bar{x}}{1+n\tau^2}\right)^2+C,$$

其中 C 与 θ 无关，易知在得到样本 x 时，θ 的后验密度为 $N\left(\frac{n\tau^2\bar{x}}{1+n\tau^2}, \frac{\tau^2}{1+n\tau^2}\right)$（因此在本例中，正态分布族也是共轭先验分布族）．由此立即得到 θ 的 Bayes 估计为

$$\delta_\tau(x)=\frac{n\tau^2\bar{x}}{(1+n\tau^2)}, \tag{3.21}$$

其风险函数不难算得为

$$R(\theta, \delta_\tau)=\frac{(n\tau^4+\theta^2)}{(1+n\tau^2)^2}, \tag{3.22}$$

δ_τ 的 Bayes 风险为

$$R_H(\delta_\tau)=\frac{\tau^2}{1+n\tau^2}. \tag{3.23}$$

由(3.21)看出，Bayes 估计 δ_τ 是通常估计 $\bar{x}$ 的一种压缩，压缩因子为 $n\tau^2/(1+n\tau^2)$．这一点是不难解释的：先验分布 $N(0, \tau^2)$ 在 0 附近的密度较大，这意味着我们先天地认为 θ 在 0 附近的可能性较大，因而将纯由样本作出的估计 $\bar{x}$ 往原点 0 方向拉一些，τ 愈小拉得愈多．另一方面，当 n 增加时，压缩因子也增加．这表明：随着抽样次数 n 的增加，样本信息的相对重要性愈来愈大，而先验信息的相对重要性则愈来愈小．又当 $\tau\to\infty$ 时（这可以解释为先验信息愈来愈少），δ_τ 趋向于通常的估计 $\bar{x}$，其风险按(3.22)则趋于估计量 $\bar{x}$ 的风险 $\frac{1}{n}$．

由于正态分布是对称单峰分布，可知即使将损失函数由 $(\theta-a)^2$ 改为 $W(|\theta-a|)$，其中 $W(t)$ 在 $t\geqslant 0$ 非降，则在先验分布 $N(0, \tau^2)$ 之下，θ 的 Bayes 估计仍为(3.21)的 δ_τ（但风险和 Bayes 风险当然有变化）．这一点的详细论证留给读者．

例 3.3 我们回忆，所谓 Gamma 分布 $G(\alpha, \beta)$，是指具密度（对 L 测度）

$$\frac{\alpha^{\beta}}{\Gamma(\beta)} x^{\beta-1} e^{-\alpha x} \quad (\text{当 } x>0, x\leqslant 0 \text{ 时为 } 0),$$

这里 $\alpha>0$, $\beta>0$ 为常数.

现设 $X=(X_1, \cdots, X_n)$, $X_1, \cdots, X_n$ 为抽自负指数分布 $G(\theta, 1)$ 的 *iid.* 样本. 令

$$g(\theta)=P_{\theta}(X_1>t_0)=e^{-\theta t_0}.$$

这可解释为元件寿命超过 t_0 的概率. 设损失函数为

$$L(\theta, a)=(g(\theta)-a)^2,$$

这表示问题为求 $g(\theta)$ 的点估计. 设先验分布为 $G(\tau^{-1}, \nu)$, $\tau>0$ 和 $\nu>0$ 都是常数.

为求 $g(\theta)$ 的 Bayes 估计，写出 $(\theta, X_1, \cdots, X_n)$ 的联合密度（对 L 测度）为

$$\theta^n e^{-\theta T_n}(\Gamma(\nu)\tau^{\nu})^{-1} e^{-\theta/\tau}\theta^{\nu-1}$$

$$(\text{当 } x_1>0, \cdots, x_n>0, \theta>0; \text{他处为 } 0).$$

此处 $T_n=\sum_{i=1}^{n} x_i$. 由此易见当得出样本 x 时，θ 的后验分布为 $G\left(T_n+\frac{1}{\tau}, n+\nu\right)$. 因而算出 $g(\theta)$ 的 Bayes 估计为

$$\frac{(T_n+\tau^{-1})^{n+\nu}}{\Gamma(n+\nu)}\int_0^{\infty} e^{-t_0 y} e^{-(T_n+\tau^{-1})y} y^{n+\nu-1} dy$$

$$=\left(1+\frac{t_0}{T_n+\tau^{-1}}\right)^{-(n+\nu)}. \tag{3.24}$$

例 3.4 设一个有限总体包含 N 个元素，其中 M 个具有性质 A. 此处 N 为已知，M 为未知，且可视为此总体的参数. 记 $\theta=M/N$. 随机地、无放回地由此总体中抽出 n 个，以 X 记其中具有性质 A 的元素个数. 为估计 θ，通常用估计量 X/n. 现设损失函数为 $(\theta-a)^2$，先验分布为

$$H(\theta=i/N)=\frac{1}{N+1} \quad (i=0, 1, \cdots, N), \tag{3.25}$$

求 θ 的 Bayes 估计.

为此需用以下的恒等式

$$\sum_{i=0}^{N-n}\binom{i+k}{k}\binom{N-k-i}{n-k}=\binom{N+1}{n+1}=\binom{N+1}{N-n}, \qquad (3.26)$$

此处 $k=0, 1, \cdots, n$. 欲证此等式,只需用

$$(1-x)^{-t}=\sum_{i=0}^{\infty}\binom{i+t-1}{t-1}x^i \quad (t=1, 2, \cdots)$$

代入恒等式

$$(1-x)^{-(n+2)}=(1-x)^{-(k+1)}(1-x)^{-(n-k+1)}$$

的两边,并比较 x^{N-n} 项的系数即可得到.

如前,分别以 P 和 P^* 表示 X 的边缘分布及(M, X)的联合分布,有

$$\begin{aligned}P(X=x)&=\sum_{i=0}^{N}H(M=i)\binom{i}{x}\binom{N-i}{n-x}\Big/\binom{N}{n}\\&=\frac{1}{N+1}\sum_{r=0}^{N-n}\binom{r+x}{x}\binom{N-r-x}{n-x}\Big/\binom{N}{n}\\&=\frac{1}{n+1}\quad(x=0, 1, \cdots, n),\end{aligned} \qquad (3.27)$$

这里最后一步用了恒等式(3.26). 又

$$P^*(M=m, X=x)=H(M=m)P^*(X=x|M=m)$$

$$=\frac{1}{N+1}\binom{m}{x}\binom{N-m}{n-x}\Big/\binom{N}{n},$$

因而

$$P^*(M=m|X=x)=\frac{n+1}{N+1}\binom{m}{x}\binom{N-m}{n-x}\Big/\binom{N}{n}$$

$$(m=0, 1, \cdots, N).$$

这个条件分布的均值即 M 的 Bayes 估计为

$$\left[\frac{n+1}{N+1}\Big/\binom{N}{n}\right]\sum_{m=0}^{N}m\binom{m}{x}\binom{N-m}{n-x}$$

$$=\left[\frac{n+1}{N+1}\Big/\binom{N}{n}\right]\sum_{m=0}^{N}(m+1)\binom{m}{x}\binom{N-m}{n-x}$$

$$-\left[\frac{n+1}{N+1}\Big/\binom{N}{n}\right]\sum_{m=0}^{N}\binom{m}{x}\binom{N-m}{n-x}\triangleq I_1-I_2, \quad (3.28)$$

令 $m-x=r$，并利用(3.26)，得

$$I_1=\left[\frac{n+1}{N+1}\Big/\binom{N}{n}\right](x+1)\sum_{m=0}^{N}\binom{m+1}{x+1}\binom{N-m}{n-x}$$

$$=\left[\frac{n+1}{N+1}\Big/\binom{N}{n}\right](x+1)\sum_{r=0}^{N-n}\binom{r+x+1}{x+1}$$

$$\times\binom{N+1-r-(x+1)}{n+1-(x+1)}$$

$$=\left[\frac{n+1}{N+1}\Big/\binom{N}{n}\right](x+1)\binom{N+2}{n+2}=\frac{N+2}{n+2}(x+1); \quad (3.29)$$

$$I_2=\left[\frac{n+1}{N+1}\Big/\binom{N}{n}\right]\sum_{r=0}^{N-n}\binom{r+x}{x}\binom{N-r-x}{n-x}$$

$$=\left[\frac{n+1}{N+1}\Big/\binom{N}{n}\right]\binom{N+1}{n+1}=1. \quad (3.30)$$

由(3.28)～(3.30)，得 M 的 Bayes 估计为

$$\frac{N+2}{n+2}(x+1)-1=\frac{N+2}{n+2}x+\frac{N-n}{n+2}.$$

而 θ 的 Bayes 估计为

$$\delta_H(x)=\frac{N+2}{N(n+2)}x+\frac{N-n}{N(n+2)}.$$

§4 共轭先验分布族

一、引言与定义

如前所述，在 Bayes 方法的基础及其使用上存在的最主要问题就是先验分布的确定．当然，比较令人信服和易为实用者所接受的选择，是基于历史资料或有根据的理论所得出的结果．除此之外，Bayes 学者还提出了一些带主观色彩的定先验分布的方法，

有的被描述成带一般性的选择法则，例如例 2.1 中提到的"同等无知"原则. 事实上，恐怕大家终究都能接受的一种看法是：尽管这些考虑在某种特定情况下可能有用，企图规定任何一种放之四海皆准的选择先验分布的法则，是徒劳无益的.

以例 2.1 中的 θ 的估计来说，可能该事件在历史上并未做过试验或者无记录，在理论上我们也对它一无所知，因而看起来，同等无知的原则及其所导致的 θ 的先验分布 $R(0,\ 1)$ 是很合理甚至是唯一可行的. 但我们也可以这样想：不仅对 θ，对 $\varphi=\theta^3$ 也是一无所知，因而我们也有理由把同等无知的原则用于 φ. 这将导致以

$$h(\theta)=\begin{cases}3\theta^2, & \text{当 } 0\leqslant\theta\leqslant 1;\\ 0, & \text{其他}.\end{cases}$$

作为 θ 的先验密度. 这样看来，没有根据认为，"同等无知"的原则是一个先天合理的原则.

但是，如果只是把 Bayes 方法当作理论论证中的一种工具，这时不存在所选先验分布是否与实际情况符合的问题，则先验分布可由人们任选，而这种选择是否恰当，对问题的处理有很大的关系. 为此目的，寻找某些一般性的选择先验分布的有效方法就是一件很有意义的事情了. 在这方面，例 3.1 中提到过的共轭先验分布族的概念是值得注意的.

我们先给一个正式的定义如下：

定义 4.1 沿用前面的记号，设 $\mathscr{H}$ 为 $\mathscr{B}_\Theta$ 上的一个概率分布族，若对任何 $H\in\mathscr{H}$ 及样本 x，可以有一个例外集 $N\in\mathscr{B}_x$, $P(N)=\int_\Theta P_\theta(N)dH(\theta)=0$)必有 $H(\cdot|x)\in\mathscr{H}$，则称 $\mathscr{H}$ 为一共轭先验分布族.

共轭先验分布族在任何情况下都是存在的. 比方说，可以取 $\mathscr{H}$ 为 $\mathscr{B}_\Theta$ 上的一切概率分布的族. 由此可见，在定义 4.1 中，并没有规定真正有意义的共轭先验分布族应是那种包含少量参数的、"自然"形式的分布族. 在下一段中，我们将讨论一个重要的例

子.

二、用充分统计量构造共轭先验分布族

设 $X_1, X_2, \cdots$ 为独立同分布变量，每个 X_i 的概率空间都是 $(\mathscr{X}, \mathscr{B}_x, \mathscr{P})$, $\mathscr{P}=(P_\Theta, \theta\in\Theta)\ll\mu$, μ 为 $\mathscr{B}_x$ 上的 σ-有限测度，记 $X_{(n)}=(X_1, \cdots, X_n)$, 则 $X_{(n)}$ 的概率空间为$(\mathscr{X}^{(n)}, \mathscr{B}_x^{(n)}, \mathscr{P}^{(n)})$, 其中

$$\mathscr{X}^{(n)}=\mathscr{X}\times\cdots\times\mathscr{X} \quad (n\text{ 重, 下同});$$

$$\mathscr{B}_x^{(n)}=\mathscr{B}_x\times\cdots\times\mathscr{B}_x,\ \mathscr{P}^{(n)}=\mathscr{P}\times\cdots\times\mathscr{P}.$$

又记
$$f_\Theta(x)=dP_\theta(x)/d\mu.$$

现作以下一些假定:

1. 对任何 n, 存在充分统计量[当然是指关于 $(\mathscr{X}^{(n)}, \mathscr{B}_x^{(n)}, \mathscr{P}^{(n)})$ 的充分统计量]

$$T_n=T_n(X_{(n)}).$$

2. 由上述假定，根据因子判别定理，应有

$$\prod_{i=1}^{n} f_\theta(x_i)=g_n(T_n(x_1, \cdots, x_n), \theta)h_n(x_1, \cdots, x_n). \tag{4.1}$$

此式本来只是对固定的 n 和 θ, 对 $x_{(n)}$ 几乎处处成立. 现在假定: 对一切 $n=1, 2, \cdots$ 及一切 $\theta\in\Theta$ 和一切 $x_{(n)}\in\mathscr{X}^{(n)}$, (4.1)式恒成立，且

$$h_n(x_1, \cdots, x_n)>0.$$

3. 存在 $\mathscr{B}_\Theta$ 上的 σ-有限测度 ν, 使对一切的 $x_{(n)}\in\mathscr{X}^{(n)}$ 及 $n=1, 2, \cdots$ 有

$$0<\int_\Theta g_n(T_n(x_1, \cdots, x_n), \theta)d\nu(\theta)<\infty. \tag{4.2}$$

引进 $\mathscr{B}_\Theta$ 上的概率分布族如下:

$$\mathscr{H}=\{H_{x_1,\cdots,x_n}: x_{(n)}\in\mathscr{X}^{(n)}, n=1, 2, \cdots\}, \tag{4.3}$$

此处

$$H_{x_1,\cdots,x_n}(B)=\int_B g_n(T_n(x_1, \cdots, x_n), \theta)d\nu(\theta) \Big/ \int_\Theta g_n(T_n(x_1, \cdots, x_n), \theta)d\nu(\theta). \tag{4.4}$$

则有

定理 4.1 在以上诸假定下，$\mathscr{H}$ 构成一个共轭先验分布族.

证 在分布族(4.3)中，任取一个

$$C(\theta)d\nu(\theta) \triangleq \frac{g_m(T_m(x_1^0, \cdots, x_m^0), \theta)}{\int_\Theta g_m(T_m(x_1^0, \cdots, x_m^0), \varphi)d\nu(\varphi)} d\nu(\theta) \quad (4.5)$$

作为 θ 的先验分布. 在得到样本 $x_1, \cdots, x_n$ 的条件下，依公式(2.5)，知 θ 的后验分布为

$$\begin{aligned} & q(x_1, \cdots, x_n, \theta)d\nu(\theta) \\ & = K_1(x_1^0, \cdots, x_m^0, x_1, \cdots, x_n)g_m(T_m(x_1^0, \cdots, x_m^0), \theta) \\ & \quad \cdot \prod_{i=1}^n f_\theta(x_i)d\nu(\theta). \end{aligned} \quad (4.6)$$

此处及下面的 $K_1, K_2, \cdots$ 都是所标出的变量的函数，其具体形式无必要写出. 当然公式(2.5)以及由之导出的(4.6)当$(x_1, \cdots, x_n)$属于某个零测集(指 $X_{(n)}$ 的边缘分布意义下的零测集)时，不必成立. 但是，既然在定义 4.1 中已容许有这么一个例外集，那么我们不妨设(4.6)总成立. 由(4.1)和(4.6)，得

$$\begin{aligned} & q(x_1, \cdots, x_n, \theta) \\ & = K_2(x_1^0, \cdots, x_m^0, x_1, \cdots, x_n)g_m(T_m(x_1^0, \cdots, x_m^0), \theta) \\ & \quad \cdot g_n(T_n(x_1, \cdots, x_n), \theta), \end{aligned} \quad (4.7)$$

又由(4.1)，有

$$\prod_{i=1}^m f_\theta(x_i^0) = g_m(T_m(x_1^0, \cdots, x_m^0), \theta)h_m(x_1^0, \cdots, x_m^0), \quad (4.8)$$

$$\prod_{i=1}^n f_\theta(x_i) = g_n(T_n(x_1, \cdots, x_n), \theta)h_n(x_1, \cdots, x_n), \quad (4.9)$$

$$\begin{aligned} & \prod_{i=1}^m f_\theta(x_i^0)\prod_{i=1}^n f_\theta(x_i) \\ & = g_{m+n}(T_{m+n}(x_1^0, \cdots, x_m^0, x_1, \cdots, x_n), \theta) \\ & \quad \cdot h_{m+n}(x_1^0, \cdots, x_m^0, x_1, \cdots, x_n). \end{aligned} \quad (4.10)$$

因为 $h_j(j=1, 2, \cdots)$ 总大于 0，由(4.8)～(4.10)，得

$$\begin{aligned} & g_m(T_m(x_1^0, \cdots, x_m^0), \theta)g_n(T_n(x_1, \cdots, x_n), \theta) \\ & = K_3(x_1^0, \cdots, x_m^0, x_1, \cdots, x_n)g_{m+n} \end{aligned}$$

$$\times (T_{m+n}(x_1^0, \cdots, x_m^0, x_1, \cdots, x_n), \theta).$$

以此代入(4.7),得

$$\begin{aligned} & q(x_1, \cdots, x_n, \theta) d\nu(\theta) \\ & = K_4(x_1^0, \cdots, x_m^0, x_1, \cdots, x_n) g_{m+n} \\ & \quad \times (T_{m+n}(x_1^0, \cdots, x_m^0, x_1, \cdots, x_n), \theta) d\nu(\theta). \quad (4.11) \end{aligned}$$

再利用条件 $\int_{\Theta} q(x_1, \cdots, x_n, \theta) d\nu(\theta) = 1$, 即得

$$\begin{aligned} & K_4(x_1^0, \cdots, x_m^0, x_1, \cdots, x_n) \\ & = \left[\int_{\Theta} g_{m+n}(T_{m+n}(x_1^0, \cdots, x_m^0, x_1, \cdots, x_n), \theta) d\nu(\theta) \right]^{-1}. \end{aligned}$$

代入(4.11)后,即知后验分布 $q(x_1, \cdots, x_n, \theta) d\nu(\theta)$ 仍在(4.3)所定义的族 $\mathscr{H}$ 内. 定理证毕.

从(4.3)及(4.4)看出,分布族 $\mathscr{H}$ 表面上依赖于"参数" $x_1, \cdots, x_n (n=1, 2, \cdots)$, 实际上只依赖于 $T_n(x_1, \cdots, x_n) (n=1, 2, \cdots)$. 因此,最有意义的情况是:不论 n 多大, $X_{(n)}$ 有一个维数固定(与 n 无关)的充分统计量 T_n. 在第一章§6中我们曾指出,这一般只在 $(P_\theta, \theta \in \Theta)$ 为指数族时才可能. 因此,可以说由本定理给出的构造共轭先验分布族的方法基本上只适用于指数族(当然,也有例外,如均匀分布族). 又由本定理给出的共轭先验分布族有时能适当扩大. 这多半是通过将 $n=1, 2, \cdots$ 扩充为 n 连续变化.

下面举几个例子说明本定理的应用.

例4.1 Bernoulli序列 设 $x_1, x_2, \cdots iid.$, 且

$$P_\theta(x_1=1) = 1 - P_\theta(x_1=0) = \theta \ (0 \leqslant \theta \leqslant 1).$$

取定理中的 $d\nu$ 为 L 测度 $d\theta$, 此处

$$T_n = \sum_1^n x_i$$

为充分统计量,而

$$\prod_{i=1}^n f_\theta(x_i) = \theta^{T_n} (1-\theta)^{n-T_n}.$$

于是由定理4.1得共轭先验分布族 $\{\beta(a, b): a, b=1, 2, \cdots\}$, 这个族可以扩大为 $\{\beta(a, b): a>0, b>0\}$.

例 4.2 Poisson 分布序列 设 $X_1, X_2, \cdots$ iid, 且

$$P_\theta(X_1=x)=e^{-\theta}\theta^x/x!,\ x=0,\ 1,\ 2,\ \cdots,\ 0<\theta<\infty.$$

取 $d\nu=d\theta$, 此处 $T_n=\sum_1^n x_i$ 为充分统计量, 而

$$\prod_{i=1}^n f_\theta(x_i)=e^{-n\theta}\theta^{T_n}/(x_1!\cdots x_n!).$$

于是由定理 4.1 得共轭先验分布族 $\{G(a,\ b):\ a,\ b=1,\ 2,\ \cdots\}$ ($G(a, b)$ 为 Gamma 分布, 见例 3.3), 这个族可以扩大为 $\{G(a, b):\ a>0,\ b>0\}$.

例 4.3 正态序列、方差已知. 设 $X_1, X_2, \cdots iid., X_1\sim N(\theta, \sigma^2)$, $-\infty<\theta<\infty$, σ^2 已知, 取 $d\nu=d\theta$. 此处

$$T_n=\frac{1}{n}\sum_1^n X_i$$

为充分统计量, 而

$$\prod_{i=1}^n f_\theta(x_i)=(\sqrt{2\pi}\,\sigma)^{-n}\exp\left(-\sum_1^n x_i^2/2\sigma^2\right)\cdot\exp\left(-\frac{n\theta^2-2nT_n\theta}{2\sigma^2}\right)$$

作为 θ 的函数看, 除去一个常数因子不计, 它正是 $N(T_n,\ \sigma^2/n)$ 的密度. 由于 T_n 可取任何值, 依定理 4.1, 得共轭先验分布族

$$\{N(a,\ \sigma^2/n):-\infty<a<\infty,\ n=1,\ 2,\ \cdots\}.$$

这个族可以扩大为

$$\{N(a,\ b^2):-\infty<a<\infty,\ b>0\}.$$

例 4.4 正态序列、均值已知. 设 $X_1,\ X_2,\ \cdots$ iid., $X_1\sim N(C,\ \sigma^2)$, C 已知而 $\sigma>0$ (参数为 σ). 取 $d\nu=d\sigma$, 此处 $T_n=\sum_1^n(X_i-C)^2$ 为充分统计量, 而

$$\prod_{i=1}^n f_\sigma(x_i)=\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n\exp(-T_n/2\sigma^2).$$

为了

$$\int_0^\infty\sigma^{-n}\exp(-T_n/2\sigma^2)d\sigma<\infty,$$

必须 $n>1$, $T_n>0$ ($T_n=0$ 的概率为 0). 于是根据定理 4.1, 得到共轭先验分布族(以对 L 测度的密度的形式表出)为

$$\left\{2a^{(n-1)/2}/\Gamma\left(\frac{n-1}{2}\right)\sigma^{-n}e^{-a/\sigma^2}: n=2,\ 3,\ \cdots,\ a>0\right\}. \quad (4.12)$$

参数的范围可扩大为 $n>1$, $a>0$ 而不丧失共轭性.

若记 $\theta=1/\sigma^2$,则易知:当 σ 有密度(4.12)时, θ 有 Gamma 分布 $G\left(a,\ \frac{n-1}{2}\right)$. 所以若以 $\theta=1/\sigma^2$ 为参数, 则共轭先验分布族为 Gamma 分布族 $\{G(a,\ b):\ a>0,\ b>0\}$. 不难证明(证明留给读者):若 $\theta=1/\sigma^2$ 的先验分布为 $G(a,\ b)$,而样本为 $x_1,\cdots,\ x_n$, 则 θ 的后验分布为 $G\left(a+\frac{1}{2}\sum_1^n(x_i-C)^2,\ b+\frac{n}{2}\right)$.

例 4.5 正态序列、均值方差都未知. 设 $X_1,\ X_2,\ \cdots iid.,\ X_1\sim N(\tilde{a},\ \sigma^2)$, $-\infty<\tilde{a}<\infty$, $\sigma>0$, $\theta=(\tilde{a},\ \sigma)$, $d\nu=d\tilde{a}d\sigma$. 此处

$$T_n=(R_n,\ S_n)=\left(\sum_1^n(X_i-\overline{X}_n)^2, \overline{X}_n\right)$$

为充分统计量$\left(\overline{X}_n=\sum_1^n X_i/n\right)$,而

$$\prod_{i=1}^n f_\theta(x_i)=(\sqrt{2\pi}\,\sigma)^{-n}\exp\left[-\frac{1}{2\sigma^2}(R_n+n(S_n-\tilde{a})^2)\right].$$

于是由定理 4.1 及例 4.4 得到 $\theta=(\tilde{a},\ \sigma)$ 的一个共轭先验分布族如下: σ 的分布为(4.12), 而在给定 σ 时, $\tilde{a}$ 的条件分布为 $N(\mu,\ \sigma^2/n)$,此处 μ 为任意的, 而 $n=1,\ 2,\ \cdots$. 此共轭分布族可扩大如下: 转化到 $\frac{1}{\sigma^2}$ 和 $\tilde{a}$. $\frac{1}{\sigma^2}$ 的分布为 $G(\alpha,\ \beta)$, $\alpha>0$, $\beta>0$. 而在给定 σ 时, $\tilde{a}$ 的条件分布为 $N(\mu,\ \sigma^2/\tau)$, μ 为任意的而 $\tau>0$. 不难证明(计算留给读者): 若$(\tilde{a},\ \sigma)$有这样一个先验分布而样本为 $x_1,\ \cdots,\ x_n$,则$(\tilde{a},\ \sigma)$的后验分布如下:

$$\frac{1}{\sigma^2}\sim G\left(\alpha+\frac{1}{2}\sum_1^n(x_i-\bar{x}_n)^2+\frac{n\tau(\bar{x}_n-\mu)^2}{2(n+\tau)},\ \beta+n/2\right),$$

而在给定 σ 的条件下, $\tilde{a}$ 的条件分布为 $N\left(\frac{n\bar{x}_n+\tau\mu}{n+\tau},\ \frac{\sigma^2}{n+\tau}\right)$.

例 4.6 多维正态、协方差阵已知. 设 $X_1,\ X_2,\ \cdots iid.,X_1\sim N_k(\theta,\ \Sigma)$, $\Sigma>0$ 已知. 留给读者验证: θ 的一个共轭先验分布族为 $\{N_k(\mu,\ \tau)\}$, 此处 μ 为任意的 k 维向量, 而 τ 为任意的 k 阶

正定阵. 而且若 θ 有先验分布 $N(\mu, \tau)$, 则在得到样本 $x_1, \cdots, x_n$ 时, θ 的后验分布为

$$N_k((\tau^{-1}+n\Sigma^{-1})^{-1}(\tau^{-1}\mu+n\Sigma^{-1}\bar{x}_n),\ (\tau^{-1}+n\Sigma^{-1})^{-1}).$$

下一例要求读者熟悉 Wishart 分布. 不熟悉这个分布的读者可以略去. 由于没有什么困难, 我们只给出结果而不作仔细计算.

例 4.7 设 $X_1, X_2, \cdots iid.$, $X_1 \sim N_k(a, \Sigma)$. 若 a 已知, 则 Σ 的一个共轭先验分布族可通过 Σ 表述如下: $\Sigma^{-1} \sim W_\alpha(k, \tau)$, 此处 $\alpha > k-1$, 而 τ 为任意 k 阶正定阵. 当 Σ^{-1} 有这样一个先验分布而得到样本 $x_1, \cdots, x_n$ 时, Σ^{-1} 的后验分布为

$$W_{\alpha+n}(k_1 [\tau^{-1}+\sum_1^n (x_i-a)(x_i-a)']^{-1}).$$

若 a 和 Σ 都未知, 则 (a, Σ) 的一个共轭先验分布族可通过 Σ 表述如下: $\Sigma^{-1} \sim W_\alpha(k, \tau)$, 其中 $\alpha > k-1$, τ 为任意 k 阶正定阵, 而在给定 Σ 的条件下, a 的条件分布为 $N_k(\mu, C^{-1}\Sigma)$, 此处 μ 为任意 k 维向量, $C>0$ 也为任意的. 当 (a, Σ) 有这样一个先验分布而样本为 $x_1, \cdots, x_n$ 时, (a, Σ) 的后验分布如下: $\Sigma^{-1} \sim W_{\alpha+n}(k, \tau^*)$, 其中

$$\tau^* = \left[\tau^{-1}+\sum_{i=1}^n (x_i-\bar{x}_n)(x_i-\bar{x}_n)' + \frac{Cn}{C+n}(\mu-\bar{x}_n)(\mu-\bar{x}_n)'\right]^{-1},$$

而在给定 Σ 时, a 的条件分布为

$$N_k\left(\frac{1}{C+n}(C\mu+n\bar{x}_n),\ \frac{1}{C+n}\Sigma\right).$$

§5 广义 Bayes 估计

一、作为 Bayes 估计的极限的广义 Bayes 估计

Bayes 估计类需要适当地扩大, 这一点从下面的简单例子可以看出来.

例 5.1 设 $X_1, \cdots, X_n$ 为抽自正态总体 $N(\theta, 1)$ 的 $iid.$ 样本 $(-\infty<\theta<\infty)$, 损失函数为 $L(\theta, a)=(\theta-a)^2$. 我们来证明: 不

论 θ 的先验分布如何，常见的估计量 $\bar{X}=\frac{1}{n}\sum_{1}^{n}x_i$ 决非 θ 的 Bayes 估计.

事实上，根据公式(2.5)，不难算出 θ 的后验分布为

$$\exp\left(-\frac{n}{2}(\theta-\bar{x})^2\right)d\xi(\theta)\Big/\int_{-\infty}^{\infty}\exp\left(-\frac{n}{2}(\theta-\bar{x})^2\right)d\xi(\theta).$$

用公式(3.8)算出 θ 的 Bayes 估计，若此 Bayes 估计为 $\bar{x}$，则应有(以 u 记 $\bar{x}$)

$$u=\int_{-\infty}^{\infty}\theta\exp\left(-\frac{n}{2}(\theta-u)^2\right)d\xi(\theta)\Big/\int_{-\infty}^{\infty}\exp\left(-\frac{n}{2}(\theta-u)^2\right)d\xi(\theta). \tag{5.1}$$

对 $u\in(-\infty,\ \infty)$ 几乎处处(L 测度)成立. 以 $h(u)$ 记 (5.1)式右边的分母，利用第一章定理 2.2，易知 $h'(u)$ 处处存在且可在积分号下求导，有

$$h'(u)=n\int_{-\infty}^{\infty}(\theta-u)\exp\left(-\frac{n}{2}(\theta-u)^2\right)d\xi(\theta).$$

再用(5.1)，知 $h'(u)=0$, a. s. L 于 $-\infty<u<\infty$. 由于 $h'(u)$ 为 u 的连续函数，故知 $h'(u)\equiv 0$ 于 $-\infty<u<\infty$，故 $h(u)$ 为一常数 C，即

$$\int_{-\infty}^{\infty}\exp\left(-\frac{n}{2}(\theta-u)^2\right)d\xi(\theta)=C\quad(-\infty<u<\infty). \tag{5.2}$$

用控制收敛定理知：当 $u\to\infty$ 时，(5.2)的左边趋于 0，故 $h(u)\equiv 0$ 于 $-\infty<u<\infty$. 由于 ξ 为一概率测度且 $\exp\left(-\frac{n}{2}(\theta-u)^2\right)>0$，这是不可能的. 这个矛盾证明了 $\bar{x}$ 确非 Bayes 估计（参看本章习题 6).

但由例 3.2 可知，尽管 $\bar{x}$ 并非 Bayes 估计，它却是一串 Bayes 估计(3.21)当 $\tau\to\infty$ 时的极限[1]. 这启示我们对 Bayes 估计的概念作一个自然的推广，即把那些本身虽不是 Bayes 估计，但可作为 Bayes 估计的极限的那些估计考虑进来. 这种估计称为“广义

1) 在此处及后面的有关讨论中，可把 τ 的取值限制为自然数.

Bayes 估计”(Generalized Bayes Estimate). 在这里必须明确所提到的极限的意义该如何规定. 概率论常用的几种收敛性, 如依概率收敛, 以概率 1 收敛等, 都可作为此处极限的意义. 但研究表明, 一个更弱一些的概念就够用了, 这就是现在我们要介绍的“正则收敛”(regular convergence)的概念.

设变量 X 的概率空间为$(\mathscr{X}, \mathscr{B}_x, P_\theta, \theta\in\Theta)$, 而行动空间$(A, \mathscr{B}_A)$为欧氏的.

定义 5.1(判决函数列的正则收敛) 设 $\delta, \delta_1, \delta_2, \cdots$ 为一串非随机化的判决函数, 若对 A 上的任何有界连续函数 u 以及 $\mathscr{X}$ 上的任何满足条件

$$\int_x |g(x)|dP_\theta(x)<\infty \quad (\text{对一切 } \theta\in\Theta) \tag{5.3}$$

的函数 g, 必有

$$\lim_{n\to\infty}\int_x u(\delta_n(x))g(x)dP_\theta(x) = \int_x u(\delta(x))g(x)dP_\theta(x) \quad (\text{对一切 } \theta\in\Theta) \tag{5.4}$$

成立. 则称$\{\delta_n\}$正则收敛于 δ, 并记为 $\delta_n\xrightarrow{r}\delta$.

若 δ、δ_n 等可以是随机化的判决函数, 则(5.4)式可相应地修改为

$$\lim_{n\to\infty}\int_x\left[\int_A u(t)\delta_n(dt|x)\right]g(x)dP_\theta(x) = \int_x\left[\int_A u(t)\delta(dt|x)\right]g(x)dP_\theta(x) \quad (\text{对一切 } \theta\in\Theta).$$

不难看出, 这包含了(5.4)作为特例. 同时在非随机化的情况下不难证明: 若对任何 $\theta\in\Theta$, 有 $\delta_n(X)\xrightarrow{p_\theta}\delta(X)$, 则 $\delta_n\xrightarrow{r}\delta$.

定义 5.2(广义 Bayes 解) 若对任何自然数 n, 判决函数 δ_n 为 Bayes 解(更确切地说, 样本空间、分布族、行动空间、损失函数等不随 n 而变, 但先验分布随 n 变化而变化), 且当 $n\to\infty$ 时, 有 $\delta_n\xrightarrow{r}\delta$, 则称 δ 为一个广义 Bayes 解.

从 Bayes 学派的观点看，引进广义 Bayes 解在实用上的意义在于，当确实缺少先验知识时，可以避免被迫作出一个有偏向的选择. 仍以例 5.1 来说：如果对 θ 的先验知识确实一无所知，则最合理的估计应是通常的估计 $\bar{x}$. 要是必须使用狭义的 Bayes 估计，就必须在缺乏根据的情况下人为地确定一个先验分布. 使用广义 Bayes 解，避免了这一点且可以说不违反 Bayes 学派的信条.

从理论的观点（不论是 Bayes 学派或否）看：引进广义 Bayes 解的意义主要在于：在很宽广的条件下，一切 Bayes 解和广义 Bayes 解构成一个完全类. 这个结论从 Bayes 学派的观点看，可以解释为 Bayes 方法的合理性. 不论怎样，这个结果提供了一个规定判决函数完全类的现实的普遍方法，不能不说有很大的理论意义.

由于涉及到过多的数学上的细节，所以我们不给出上述结果（即 Bayes 和广义 Bayes 解的类的完全性）的严格表述与证明了. 读者可以参看 Wald[9]. 事实上，所述结果是 Wald 这本经典著作的一个主要结果，也可以参看 Farrell[10]. 1976 年，Brown 在一份未公开发表的材料中对这个问题作了透彻的论述.

二、通过广义先验分布定义广义 Bayes 解

仍回到例 5.1，根据例 3.2，估计量 $\bar{x}$ 是一串 Bayes 估计 (3.21) 当 $\tau\to\infty$ 时的极限. 如果从先验分布着眼，则当 $\tau\to\infty$ 时，大致可以说，先验分布 $N(0, \tau^2)$ 趋向于 $(-\infty, \infty)$ 上的"均匀分布"，即 L 测度. 但 L 测度不是一个概率测度，不能直接取为先验分布. 问题就在这里. 但以上的考虑启示我们：估计量 $\bar{x}$ 可能会与作为分布 $N(0, \tau^2)$ 的极限的 L 测度发生某种联系，而这种情况可以作为通过推广先验分布的概念来推广 Bayes 解的依据.

仔细考察一下 Bayes 解的定义 2.2，可以看出在此定义中，H 为概率测度并非必需. 因此从形式上说，我们可以把 $\mathscr{B}_\Theta$ 上的任一测度 H 称为一个"先验分布"，而仍按定义 2.2 来定义 Bayes 解. 问题在于，对先验分布作了这种推广以后，"后验风险最小"的

原则能否维持？细察公式(2.4)，可以看出：要由(2.4)定义的$H(\cdot|x)$仍构成一个概率分布，只要

$$0<\int_{\Theta} f(x, \theta) dH(\theta)<\infty \tag{5.5}$$

就行，而不必要有$H(\Theta)=1$. 这些考虑引导到如下的定义：

定义 5.3(广义先验分布) 设X的概率空间为$(\mathscr{X}, \mathscr{B}_x, P_\theta, \theta\in\Theta)$，对任何$\theta\in\Theta$, $P_\theta\ll\mu$, μ为$\mathscr{B}_x$上的σ-有限测度，记

$$dP_\theta(x)/d\mu=f(x, \theta).$$

设H为$\mathscr{B}_\Theta$的一个测度，满足条件$H(\Theta)=\infty$及(5.5)(对一切x，或可以有例外集$N\in\mathscr{B}_x$，使$\int_\Theta P_\theta(N)dH(\theta)=0$)，则称$H$为一个广义先验分布(Generalized Prior Distribution).

条件$H(\Theta)=\infty$的理由是：若$H(\Theta)=C$ $(0<C<\infty)$，则用H/C代H，就回到通常先验分布的情况.

定义 5.4 沿用前面的记号. 若H为一个广义先验分布，则(2.1)定义的$R_H(\delta)$称为判决函数δ的广义 Bayes 风险. 使广义 Bayes 风险达到最小(在指定的判决函数类中)的判决函数δ_H称为相应于此广义先验分布H的广义 Bayes 解.

不难验证：在定义 5.3 和 5.4 之下，就例 5.1 而言，L测度为一个广义先验分布，而相应的广义 Bayes 解即为通常的估计量$\bar{x}$.

下面证明：在广义先验分布下，"后验风险最小"的原则仍适用.

定理 5.1 若判决函数δ满足条件

$$\int_\Theta L(\theta, \delta(x))H(d\theta|x) =\inf\left\{\int_\Theta L(\theta, a)H(d\theta|x):\ a\in A\right\}, \tag{5.6}$$

这里$H(\cdot|x)$由(2.4)给出(又(5.6)对x可有一个例外集$N\in\mathscr{B}_x$满足条件$\int_\Theta P_\theta(N)dH(\theta)=0$)，则$\delta$就是在广义先验分布$H$之下的广义 Bayes 解.

证明很简单. 只需注意对任何判决函数δ^*,有

$$R_H(\delta^*)=\int_\Theta\left[\int_x L(\theta,\ \delta^*(x))f(x,\ \theta)d\mu(x)\right]dH(\theta)$$

$$=\int_x\left[\int_\Theta L(\theta,\ \delta^*(x))H(d\theta|x)\right]t(x)d\mu(x).$$

此处$t(x)=\int_\Theta f(x,\ \theta)dH(\theta)$.

以上提出了广义 Bayes 解的两个定义. 自然地会产生在何种情况下二者一致的问题. 关于这个问题可参看[11] 和 [12]. 在[11]中, Sack 仔细讨论了单参数指数族的情况. Sacks 的研究证明了在这种情况下,前述两个广义 Bayes 解的定义是一致的(加上某些条件). 这类工作的仔细叙述涉及不少复杂的细节, 此处从略了.

现在考察一个例子.

例 5.2 再取例 5.1 并取广义先验分布为L测度$d\theta$. 易见当得出样本$x_1,\ \cdots,\ x_n$时, 依 (2.4) 算出的θ的"后验分布"为$N\left(\bar{x},\ \frac{1}{n}\right)$. 设损失函数有$L(\theta,\ a)=W(|\theta-a|)$的形状, 则依定理 5.1, 广义 Bayes 解δ应满足条件

$$\int_{-\infty}^{\infty}\exp\left(-\frac{n}{2}(\theta-\bar{x})^2\right)W(|\theta-\delta(x)|)d\theta$$

$$=\inf\left\{\int_{-\infty}^{\infty}\exp\left(-\frac{n}{2}(\theta-\bar{x})^2\right)W(|\theta-a|)d\theta:\right.$$

$$\left.-\infty<a<\infty\right\}.$$

下面证明:只要$W(t)$在$t\geqslant 0$处单调非降, 则$\delta(x)=\bar{x}$满足这个条件,因而为广义 Bayes 估计.

不难验证,要证明这个事实,只需证明:函数

$$g(a)=\int_{-\infty}^{\infty}W(|\theta+a|)f(|\theta|)d\theta\quad(-\infty<a<\infty)$$

在$a=0$处达到最小值. 此处$f(t)$在$t\geqslant 0$处非增非负. 显然,

$$g(-a)=g(a).$$

故不失普遍性,可设$a>0$, 有

$$
\begin{aligned}
g(a)-g(0)=&\int_0^{\infty}[W(|\theta+a|)-W(|\theta|)]f(|\theta|)d\theta \\
&+\int_{-a/2}^{0}[W(|\theta+a|)-W(|\theta|)]f(|\theta|)d\theta \\
&+\int_{-\infty}^{-a/2}[W(|\theta+a|)-W(|\theta|)]f(|\theta|)d\theta \\
\triangleq & I_1+I_2+I_3.
\end{aligned}
\tag{5.7}
$$

在积分 I_3 中作代换 $\theta=-t$, 注意 $|-t+a|=|t-a|$, $|-t|=|t|$, 再作代换 $t=\theta+a$, 即得

$$
I_3=-\int_{-a/2}^{\infty}[W(|\theta+a|)-W(|\theta|)]f(|\theta+a|)d\theta.
$$

于是

$$
\begin{aligned}
&I_1+I_3 \\
&=\int_0^{\infty}[W(|\theta+a|)-W(|\theta|)][f(|\theta|)-f(|\theta+a|)]d\theta \\
&\quad-\int_{-a/2}^{0}[W(|\theta+a|)-W(|\theta|)]f(|\theta+a|)d\theta \\
&\triangleq I_4+I_5.
\end{aligned}
\tag{5.8}
$$

由于 $a>0$ 以及对 W 和 f 的单调性假设知: 在 $\theta>0$ 时, 积分 I_4 的被积函数非负, 故 $I_4\geqslant 0$. 因而, 欲证 $g(a)-g(0)\geqslant 0$, 只需证 $I_2+I_5\geqslant 0$. 但

$$
\begin{aligned}
&I_2+I_5 \\
&=\int_{-a/2}^{0}[W(|\theta+a|)-W(|\theta|)][f(|\theta|)-f(|\theta+a|)]d\theta.
\end{aligned}
\tag{5.9}
$$

因为在 $-\frac{a}{2}<\theta<0$ 内有 $|\theta+a|\geqslant|\theta|$, 由对 W 和 f 的单调性假定知: (5.9)右边积分的被积函数在 $-\frac{a}{2}<\theta<0$ 内非负 这就证明了所要的结果.

三、广义 Bayes 区间估计

考察一下广义 Bayes 解的概念用于区间估计的情况, 也许有助于加深对这个问题的理解.

如在例 2.1 中描述过的，Bayes 区间估计一种确定方法如下：设先验分布为 H. 指定 $\alpha\in(0,1)$. 有了样本 x 以后，取区间 $[A(x), B(x)]$，使

$$H(A(x)\leqslant\theta\leqslant B(x)\,|\,x)=1-\alpha. \tag{5.10}$$

这种作法的好处是区间 $[A(x), B(x)]$ 有指定的"事后信度"$1-\alpha$（我们前已指出，这不等价于 Neyman 理论中的置信系数 $1-\alpha$）. 在满足条件(5.10)的前提下，可以对 $[A(x), B(x)]$ 提出某种优良性要求，如其长最小的类. 现在我们来举例证明：例 5.1 的现象在区间估计中也有，即最常用的区间估计可以不是 Bayes 区间估计.

例 5.3 设 $X_1, \cdots, X_n$ 为抽自 $N(\theta, 1)$ 的 iid. 样本. 这时，常用的区间估计是 $[\bar{x}-b/\sqrt{n}, \bar{x}+b/\sqrt{n}]$，$b$ 由

$$\frac{1}{\sqrt{2\pi}}\int_{-b}^{b}e^{-t^2/2}dt=1-\alpha$$

确定，现证：不论先验分布 H 如何取，上述区间估计决非在(5.10)意义下的 Bayes 区间估计.

因 $\bar{x}$ 为充分统计量且仍为正态分布，不失普遍性，可设 $n=1$. 在得到样本 x 时，θ 的后验分布为

$$C(x)\exp\left(-\frac{1}{2}(x-\theta)^2\right)dH(\theta),$$

此处
$$C(x)=\left[\int_{-\infty}^{\infty}\exp\left(-\frac{1}{2}(x-\theta)^2\right)dH(\theta)\right]^{-1}.$$

因此，若 $x\pm b$ 为(5.10)意义下的 Bayes 估计，应有

$$C(x)\int_{x-b}^{x+b}\exp\left(-\frac{1}{2}(x-\theta)^2\right)dH(\theta)=1-\alpha,$$

即

$$\int_{x-b}^{x+b}\frac{1}{A}\exp\left(-\frac{1}{2}(x-\theta)^2\right)dH(\theta)$$
$$=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}(x-\theta)^2\right)dH(\theta) \tag{5.11}$$

对 $-\infty<x<\infty$ 成立，此处

$$A=\int_{-b}^{b}e^{-t^2/2}dt.$$

现在引进三个独立随机变量 ξ_1、ξ_2、ξ_3 如下：ξ_1 的分布为 H，$\xi_2 \sim N(0, 1)$，而 ξ_3 有密度(对 L 测度)

$$g(t)=\begin{cases} e^{-t^2/2}/A, \text{当} |t| \leqslant b, \\ 0, \text{当} |t|>b. \end{cases}$$

由 (5.11) 可知，$\xi_1+\xi_2$ 与 $\xi_1+\xi_3$ 有同一的密度函数. 故若以 $f_i(t)$ $(i=1, 2, 3)$ 分别记 $\xi_i (i=1, 2, 3)$ 的特征函数，将有

$$f_1(t) f_2(t)=f_1(t) f_3(t).$$

由于在 $t=0$ 的邻域内 $f_1(t) \neq 0$，知 $f_2(t)=f_3(t)$ (当 $|t|$ 充分小). 但 $f_2(t)=e^{-t^2/2}$，因而 $f_3(t)=e^{-t^2/2}$ (当 $|t|$ 充分小). 由特征函数与矩的关系将得 $E(\xi_3^2)=1$. 但

$$f(b) \triangleq E(\xi_3^2)=\int_0^b e^{-t^2/2} t^2 \, dt \Big/ \int_0^b e^{-t^2/2} \, dt$$

有 $\lim\limits_{b \to \infty} f(b)=1$，且

$$f'(b)=\left[b^2 e^{-b^2/2} \int_0^b e^{-t^2/2} dt-e^{-b^2/2} \int_0^b t^2 e^{-t^2/2} dt\right] \Big/ \left(\int_0^b e^{-t^2/2} \, dt\right)^2>0.$$

因而 f 为增函数，故对于一切 $b>0$，有 $f(b)<1$. 就是说，对于 $b>0$，有 $E(\xi_3^2)<1$. 这个矛盾证明了所要的结果.

但是不难验证：$x \pm b$ 可以作为一串 Bayes 区间估计[在(5.10)的意义下，且要求 $B(x)-A(x)$ 最小]的极限. 为此，只需取先验分布 $N(0, \tau^2)$，再令 $\tau \to \infty$ 即可. 对其仔细的验证留给读者去作. 又若取广义先验分布为 $d\theta$，且仍按(2.4)式计算后验密度，则 $x \pm b$ 正好是这个广义先验分布之下的广义 Bayes 区间估计. 下例中有同样的现象.

例 5.4 设 $X_1, \cdots, X_n$ 为抽自 $N(0, \sigma^2)$ $(\sigma>0)$ 的 *iid.* 样本，取 σ 的广义先验分布为 $d\sigma/\sigma$，则通过公式(2.4)算出的 σ 的后验分布为 Gamma 分布 $G\left(y, \dfrac{n}{2}\right)$，此处 $y=\dfrac{1}{2} \sum\limits_1^n x_i^2$，以 $g(\cdot \mid y, n)$ 记这个分布的密度函数，则由 g 的单峰性可知，满足条件(5.10)的最短的区间估计 $[A(y), B(y)]$ 由方程组

$$\begin{cases} g(A(y)|y, n) = g(B(y)|y, n), \\ \int_{A(y)}^{B(y)} g(t|y, n)dt = 1-\alpha. \end{cases} \tag{5.12}$$

定出.

如果对 $\frac{1}{\sigma^2}$ 给以正常的先验分布 $G\left(\frac{1}{m}, \frac{1}{m}\right)$ $(m=1, 2, \cdots)$，算出在先验分布 $G\left(\frac{1}{m}, \frac{1}{m}\right)$之下满足条件(5.10)的最短区间估计 $[A_m(x), B_m(x)]$，则当 $m\to\infty$ 时，将有 $A_m\to A$, $B_m\to B$. A、B 是由(5.12)定出的. 因此在本例中，广义 Bayes 解的两个意义也一致. 对上述事实的验证留给读者完成.

例 5.5 设 $X_1, \cdots, X_n$ 为抽自 $N(a, \sigma^2)$的 $iid.$ 样本，a和σ都未知(a 任意, $\sigma>0$). 在 Neyman 理论中，通常 a 和 σ^2 的区间估计是基于 t-分布和 χ^2-分布. 下面证明：若取(a, σ)的广义先验分布为 $\frac{1}{\sigma}da\,d\sigma$，则上述通常的区间估计是广义 Bayes 区间估计.

按(2.4)，在得到样本 $x_1, \cdots, x_n$ 时，(a, σ)的后验密度(对 L 测度，且只在 $\sigma>0$ 处)为

$$C_n(x_1, \cdots, x_n)\sigma^{-(n+1)}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^n(x_i-a)^2\right]da\,d\sigma. \tag{5.13}$$

此处 $C_n(x_1, \cdots, x_n)$及以下的 D_n、E_n 等都与 a、σ 无关. 由(5.13)算出 a 的边缘密度，也可以算作是 a 本身的后验密度，为

$$\begin{aligned} &D_n(x_1, \cdots, x_n)\left[\sum_1^n (x_i-a)^2\right]^{-n/2} \\ &\quad = E_n(x_1, \cdots, x_n)\,[1+n(\bar{x}-a)^2/s^2]^{-n/2}, \end{aligned}$$

此处 $s^2=\sum_1^n(x_i-\bar{x})^2$. 由此易见，若令

$$a^* = \sqrt{n(n-1)}\,(a-\bar{x})/s, \tag{5.14}$$

则 a^* 的分布为自由度 $n-1$ 的中心 t-分布，于是由 t-分布密度的对称单峰性即得满足条件(5.10)且最短的区间估计就是通常的 t-区间估计.

另一方面，将(5.13)对 a 在$(-\infty, \infty)$积分，易得σ的后验密度(在 $\sigma>0$处)为

$$F_n(x_1, \cdots, x_n)\sigma^{-n}\exp(-s^2/2\sigma^2). \tag{5.15}$$

由此易得 $s^2/\sigma^2 \sim \chi^2_{n-1}$（应注意的是：此式虽在形式上与初等统计中见到的一样，但实际含义却完全不同：在初等统计中是 s^2 为随机变量而 σ^2 固定；在此则 s^2 为常数而 σ 为随机变量，有分布(5.15)）. 由 χ^2 密度的单峰性可知，适合条件 (5.10) 且长度最短的、σ^2 的 Bayes 区间估计有形式$[s^2/d, s^2/c]$，c、d 选择后，使在约束

$$\int_c^d \chi^2_{n-1}(u)du = 1-\alpha \quad (\chi^2_{n-1}(\cdot)\text{为 } \chi^2_{n-1} \text{ 的密度}) \tag{5.16}$$

之下，$\dfrac{1}{c}-\dfrac{1}{d}$ 达到最小. 不难验证：c 和 d 由(5.16)以及

$$\chi^2_{n-1}(c)c^2 = \chi^2_{n-1}(d)d^2 \tag{5.17}$$

联立解出. 这需要用 χ^2–分布表并用试探法.

不难证明：本例中的 Bayes 区间估计也可以作为正常先验分布下的 Bayes 区间估计的极限. 对这一点的详细验证留给读者完成.

以上诸例就正态总体 $N(a, \sigma^2)$ 的参数提出了三个重要的广义先验分布：

$$dH = da, \quad \text{当 } \sigma^2 \text{ 已知;}$$

$$dH = d\sigma/\sigma, \quad \text{当 } a \text{ 已知;}$$

$$dH = da\,d\sigma/\sigma, \quad \text{当 } a, \sigma \text{ 都未知.}$$

这几个广义先验分布在许多理论问题中很有用. 其更深的根据如下述：若在空间 $\{-\infty<a<\infty\}$、$\{\sigma>0\}$ 和 $\{-\infty<a<\infty, \sigma>0\}$ 分别引进变换群

$$a' = a+c \quad (-\infty<c<\infty);$$

$$\sigma' = c\sigma \quad (c>0);$$

$$\begin{cases} a' = a+c_1 \\ \sigma' = c_2\sigma \end{cases} \quad (-\infty<c_1<\infty, c_2>0).$$

则上述三个测度分别是在这三个变换群之下的不变测度，称为 Haar 测度（见[13]）.

§6 经验Bayes估计

一、引言与定义

经验 Bayes 方法 (Empirical Bayes Method) 是 H. Robbins 在 1955 年提出的(见[14]). 他提出的这种方法的思想受到统计学者相当的重视. 统计学界的元老 J. Neyman 甚至称它为统计判决的"两大突破之一". 二十多年来, 不少统计学者将 Robbins 方法的思想用于种种统计问题, 得出了不少结果. 1970 年出版了这方面的专著[15]. Robbins 的文章[16]是这个方法的一个很好的介绍.

前已多次指出, Bayes 方法最大的问题, 就在于要求参数 θ 有一定的先验分布. 即使在某项具体问题中这个要求可以认为是合理的, θ 的先验分布一般也无法确知, 因而不能不对它作一种人为的假设性的规定. 因为当先验分布的指定与实际情况不符时, 所得的解会受到较大的影响, 这样一来, 在对先验分布无法基本确定时, Bayes 方法的适用性和优越性就颇值得怀疑了. 经验 Bayes 方法的提出就是针对这个问题的.

从本质上说, 经验 Bayes 方法无非就是在前面提到过的、通过历史资料(即"经验")以决定先验分布的方法. 这个方法使用的前提是: 参数 θ 确是通常意义下的随机变量, 其分布 H 在历史资料积累的过程中一直维持不变. 在本章开头处提到的那个估计废品率的问题, 是适合这个条件的典型例子.

下面我们来说明经验 Bayes 方法的基本思想以及经验 Bayes 解的严格定义.

设变量 X 的概率空间为 $(\mathscr{X}, \mathscr{B}_x, P_\theta, \theta\in\Theta)$, 行动空间和损失函数分别为 $(A, \mathscr{B}_A)$ 和 $L(\theta, a)$. 假定我们在过去已多次面对过这个统计判决问题. 在第 i 次碰到这个问题时, 样本为 x_i, 参数真值为 θ_i(以上面提到的废品率的例子来说, θ_i 是第 i 个试验日的产品的废品率, x_i 是当日所抽的 m 个产品的废品数). 我们假定 θ 有

一定的先验分布 H，且只知 H 属于某个分布族 $\mathscr{H}$，而 $\theta_1, \cdots, \theta_n$ 可以看成是从分布 H 中抽出的 iid."样本"．事实上，我们当然并不知道 $\theta_1, \cdots, \theta_n$ 的确定值，而只知道 $x_1, \cdots, x_n$，它们是独立同分布，且每一个的分布就是由(2.3)式决定的 P．因为分布 P 与先验分布 H 有关，在样本 $x_1, \cdots, x_n$ 中也就包含了 H 的信息，n 愈大，所包含的信息应当愈多．

现在再一次面对上述统计判决问题，得到的样本是 x，真参数值为 θ．在求 Bayes 解时，可以参考由历史资料 $x_1, \cdots, x_n$ 中获得的关于 H 的信息，以选定一个判决函数 δ．这个 δ 将与 $x_1, \cdots, x_n$ 有关，因而可记为 $\delta_n(x|x_1, \cdots, x_n)$．我们希望它的 Bayes 风险接近真正的 Bayes 解 $\delta_H(x)$ 的 Bayes 风险 $R_H(\delta_H)$，并且当 $n\to\infty$ 时，趋于 $R_H(\delta_H)$ 为极限．但 $\delta_n(x|x_1, \cdots, x_n)$ 该如何计算？首先，固定 $x_1, \cdots, x_n$，这时 $\delta_n(x|x_1, \cdots, x_n)$ 只与 x 有关，其 Bayes 风险为(P^* 为 (X, θ) 的联合分布，E^* 表示在这个分布下求期望)：

$$
\begin{aligned}
R_H(\delta_n|x_1, \cdots, x_n) &= E^*[L(\theta, \delta_n(X|x_1, \cdots, x_n))] \\
&= \int_{\mathscr{X}\times\Theta} L(\theta, \delta_n(x|x_1, \cdots, x_n))dP^*(x, \theta) \\
&= \int_{\Theta}\left[\int_{\mathscr{X}} L(\theta, \delta_n(x|x_1, \cdots, x_n))dP_\theta(x)\right]dH(\theta).
\end{aligned} \tag{6.1}
$$

由于 $x_1, \cdots, x_n$ 也是随机的，还要对它们求一次期望，这样得到这种形式的 δ_n 的"全面"Bayes 风险为

$$
\begin{aligned}
R_H^*(\delta_n) &= E_p[R_H(\delta_n|x_1, \cdots, x_n)] \\
&= \int_{\mathscr{X}\times\cdots\times\mathscr{X}} R_H(\delta_n|x_1, \cdots, x_n)dP(x_1)\cdots dP(x_n).
\end{aligned} \tag{6.2}
$$

由于对任何 $x_1, \cdots, x_n$ 有 $R_H(\delta_n|x_1, \cdots, x_n) \geqslant R_H(\delta_H)$，必有

$$R_H^*(\delta_H) \geqslant R_H(\delta_H).$$

就是说，任何经验 Bayes 判决函数 δ_n 的全面 Bayes 风险，都不能低于 Bayes 解的 Bayes 风险，这引出如下的定义：

定义 6.1 任何同时依赖于历史样本 $x_1, \cdots, x_n$ 和当前样本 x 的判决函数 $\delta_n = \delta_n(x|x_1, \cdots, x_n)$ 称为经验 Bayes 判决函数，如果对任何 $H \in \mathscr{H}$，有

$$\lim_{n\to\infty} R_H^*(\delta_n) = R_H(\delta_H). \tag{6.3}$$

则称 δ_n 为渐近最优(简记为 a. o.)的经验 Bayes 判决函数.

应当注意的是: 在经验 Bayes 判决函数 $\delta_n(x|x_1, \cdots, x_n)$ 中, 历史样本 $x_1, \cdots, x_n$ 与当前样本 x 的作用是不一样的. $x_1, \cdots, x_n$ 的作用在于由之获得关于先验分布 H 的信息以帮助选定一个尽可能接近真正的 Bayes 解 $\delta_H(x)$ 的判决函数 $\delta_n^*(\cdot|x_1, \cdots, x_n)$, 而推断当前参数值 θ 的任务则落在当前样本 x 的头上. 例如估计当日废品率 θ 的任务落在当日抽样结果 x 的头上, 而过去积累的资料 $x_1, \cdots, x_n$ 则用来选定一个适当的估计量.

在上面我们只是说, 历史样本 $x_1, \cdots, x_n$ 中包含了先验分布 H 的信息. 我们并无任何把握断言, 即使当 $n\to\infty$ 时, 这种信息能充足到足以确定先验分布 H, 因而也就没有把握断言 a. o. 经验 Bayes 解必存在. 事实上, 即使像二项分布这么简单的情况, 如果可能的先验分布 H 的族 $\mathscr{H}$ 不满足一种很特殊的条件, 则 a. o. 经验 Bayes 解就不会存在. 因此, 为了 a. o. 经验 Bayes 解存在, 常有必要将先验分布族 $\mathscr{H}$ 加以相当的人为的限制, 有时甚至要限制 $\mathscr{H}$ 是一个只包含少数参数的分布族. 但这样一来, 就与我们引进经验 Bayes 方法(这个方法的目的就在于克服规定先验分布的人为性) 的初衷有所背离. 最后, 在 $\mathscr{H}$ 比较复杂的情况, 即使 a. o. 经验 Bayes 解存在, (6.3)式的收敛速度也不会很快.

所以我们认为, Neyman 将经验 Bayes 估计的引进视为统计判决理论的一个"突破"的看法(见[18]), 在估价上似乎失之过高. 当然, 也许 Neyman 的着眼点主要不在于这个方法的实际应用方面, 而在于它的理论意义, 即把古典统计推断法与 Bayes 方法在一定程度上沟通起来.

下面举一个简单的例子:

例 6.1 设 $X\sim N(\theta, 1)$, 损失函数 $L(\theta, a)=(\theta-a)^2$, θ 的先验分布只知道属于族 $\mathscr{H}=\{N(0, \sigma^2), \sigma>0\}$. 设 $x_1, \cdots, x_n$ 为历史样本, 由于 X 在 θ 的先验分布 $N(0, \sigma^2)$ 之下的边缘分布为

$N(0,\ 1+\sigma^2)$(这一事实的验证留给读者)，得 σ^2 的估计为

$$\hat{\sigma}_n^2=\frac{1}{n}\sum_1^n x_i^2-1. \tag{6.4}$$

设当前样本为 x. 取 θ 的先验分布为 $N(0,\ \hat{\sigma}_n^2)$，由例 3.2 知，θ 在这个先验分布下的 Bayes 估计为

$$\delta_n(x|x_1,\ \cdots,\ x_n)=\frac{\hat{\sigma}_n^2}{1+\hat{\sigma}_n^2}x=\frac{\sum_1^n x_i^2-1}{\sum_1^n x_i^2}x, \tag{6.5}$$

其 Bayes 风险(在 $x_1,\ \cdots,\ x_n$ 固定的条件下)由在(3.23)式中令 $n=1$ 和 $\tau^2=\hat{\sigma}_n^2$ 而得到，为

$$R_H(\delta_n|x_1,\ \cdots,\ x_n)=\hat{\sigma}_n^2/(1+\hat{\sigma}_n^2),$$

因而得到 δ_n 的无条件 Bayes 风险为

$$R_H^*(\delta_n)=E[\hat{\sigma}_n^2/(1+\hat{\sigma}_n^2)]. \tag{6.6}$$

由大数定律，以概率为 1 地成立

$$\hat{\sigma}_n^2\to(1+\sigma^2)-1=\sigma^2.$$

于是根据(6.6)，由控制收敛定理，得

$$\lim_{n\to\infty}R_H^*(\delta_n)=\frac{\sigma^2}{1+\sigma^2}.$$

但这就是当 σ^2 已知时，在 θ 的先验分布为 $N(0,\ \sigma^2)$ 时，θ 的 Bayes 估计的 Bayes 风险. 因而(6.3)成立，这证明了(6.5)确定的 δ_n 是相对于先验分布族 $\{N(0,\ \sigma^2):\sigma>0\}$ 的 a. o. 经验 Bayes 估计.

二、Poisson 分布参数的经验 Bayes 估计和其他例子

本段我们仔细讨论 Poisson 分布

$$P_\theta(X=x)=\frac{e^{-\theta}}{x!}\theta^x\quad(x=0,\ 1,\ 2,\ \cdots;\ \theta\geqslant 0) \tag{6.7}$$

的参数 θ 在平方损失 $L(\theta,\ a)=(\theta-a)^2$ 之下的经验 Bayes 估计. 由这个例子可看出：即使在不太复杂的情况下，要证明某个看来自然的估计是 a. o. 经验 Bayes 估计，也是很费事的. 这个例子有意义之处在于，它对先验分布族的限制甚少.

设先验分布为 H，则

$$P(X=x)=f_H(x)=\int_0^\infty \frac{e^{-\theta}}{x!}\theta^x dH(\theta) \quad (x=0,\ 1,\ \cdots) \quad (6.8)$$

是 X 的边缘分布．在平方损失下，依公式(3.8)，不难算出对先验分布 H 的 Bayes 估计为

$$\delta_H(x)=(x+1)f_H(x+1)/f_H(x). \quad (6.9)$$

若 H 未知，但有了历史样本 $x_1,\ \cdots,\ x_n$，它们是从分布(6.8)中抽出的 *iid*. 样本，故由之可作出 $f_H(x)$ 的估计．为了避免(6.9)的分母为 0，取这个估计为

$$u_n(x,\ x_1,\ \cdots,\ x_n)=\frac{1}{n+1}\{(x_1,\ \cdots,\ x_n \text{ 中等于 } x \text{ 的个数})+1\} \quad (6.10)$$

以此代替(6.9)的 $f_H(x)$［以及以 $u_n(x+1,\ x_1\cdots,x_n)$代 $f_H(x+1)$］，得 θ 的经验 Bayes 估计为

$$\delta_n(x\,|\,x_1,\ \cdots,\ x_n)$$
$$=(x+1)u_n(x+1,\ x_1,\ \cdots,\ x_n)/u_n(x,\ x_1,\ \cdots,\ x_n). \quad (6.11)$$

我们的目标是证明：在分布族

$$\mathscr{H}=\left\{H:\int_0^\infty \theta^2\, dH(\theta)<\infty\right\} \quad (6.12)$$

之下，(6.11)为 θ 的 a. o. 经验 Bayes 估计．据 Robbins 在[16]中提到，Johns 在其博士论文[19]中曾证明了这个事实，但 Johns 的工作迄未公开发表．下面我们自己拟了一个证明：

首先，根据定义 6.1，有

$$R_H^*(\delta_n)=E_{(X,\theta)}^*[E_{(x_1,\cdots,x_n)}(\delta_n(X\,|\,x_1,\ \cdots,\ x_n)-\theta)^2]. \quad (6.13)$$

这里 $E_{(x_1,\cdots,x_n)}$ 的意思是：在求期望时，x 看作常数，$x_1,\ \cdots,\ x_n$ 为 *iid*.，每一个有分布(6.8)．$E_{(X,\theta)}^*$ 的意思是：期望值对$(X,\ \theta)$的联合分布为 P^*(见(2.2)式)去取．

为了证明 δ_n 是相对于 $\mathscr{H}$ 的 a. o. 经验 Bayes 估计，只需证明以下三条：

a. $\lim\limits_{n\to\infty} E_{(x_1,\cdots,x_n)}(\delta_n(X\,|\,x_1,\ \cdots,\ x_n)-\theta)^2=(\delta_H(x)-\theta)^2$;

b. 存在只依赖于 x 和 θ 的函数 $G(x,\ \theta)$，使

$$E_{(x_1,\cdots,x_n)}[\delta_n(X|x_1, \cdots, x_n)-\theta]^2 \leqslant G(x, \theta);$$

c. $E^*[G(x, \theta)]=\int_{\mathscr{X}\times\Theta} G(x, \theta)dP^*(x, \theta)<\infty$

对任何 $H\in\mathscr{H}$(注意 G 与 H 无关)。事实上,若这三条已证明,则由(6.13),用控制收敛定理,得

$$\lim_{n\to\infty} R_H^*(\delta_n)=E_{(X,\theta)}^*[(\delta_H(x)-\theta)^2]=R_H(\delta_H), \quad H\in\mathscr{H},$$

而这就证明了 δ_n 为 a. o. 经验 Bayes 解.

为了证明 a～c, 需要以下两条引理:

引理 6.1 设(ξ_1, ξ_2, ξ_3)服从多项分布 $M(p_1, p_2, p_3)$:

$$P(\xi_i=n_i, i=1, 2, 3)=\frac{n!}{n_1!n_2!n_3!}p_1^{n_1}p_2^{n_2}p_3^{n_3},$$

此处 $p_i>0$, $n_i(i=1, 2, 3)$为非负整数, $\sum_1^3 p_i=1$, $\sum_1^3 n_i=n$. 则有

$$E\left(\frac{1+\xi_2}{1+\xi_1}\right)^2 \leqslant 16\left(\frac{p_2}{p_1}\right)^2+4. \tag{6.14}$$

证 有

$$E\left(\frac{1+\xi_2}{1+\xi_1}\right)^2=\sum{}'\frac{n!}{n_1!n_2!n_3!}p_1^{n_1}p_2^{n_2}p_3^{n_3}\left(\frac{1+n_2}{1+n_1}\right)^2. \tag{6.15}$$

此处 $\sum'$ 表示对 n_1, n_2, n_3 在上述范围内求和. 由于 $\frac{2+n_1}{1+n_1}\leqslant 2$, 又

$$\frac{(1+n_2)^2}{n_2!}\leqslant\frac{8}{(n_2-2)!} \quad (当 n_2\geqslant 2);$$

$$\frac{2+n_1}{1+n_1}\leqslant 2 \quad (当 n_1=0, 1, 2, \cdots).$$

以 $\sum''$ 表示对 n_1, n_2, n_3 在 $\sum'$ 的范围且 $n_2\geqslant 2$ 求和, 有

$$E\left(\frac{1+\xi_2}{1+\xi_1}\right)^2$$

$$\leqslant 16\sum{}''\frac{n!}{(n_1+2)!(n_2-2)!n_3!}p_1^{n_1+2}p_2^{n_2-2}p_3^{n_3}\left(\frac{p_2}{p_1}\right)^2$$

$$+4\sum{}'\frac{n!}{n_1!n_2!n_3!}p_1^{n_1}p_2^{n_2}p_3^{n_3}$$

$$\leqslant 16\left(\frac{p_2}{p_1}\right)^2+4.$$

这就证明了(6.14). 注意此式的特点是:左边与n有关,而右边与n无关.

下一引理的证明参看[19].

引理 6.2 设ξ服从二项分布$B(n, p)$, $\varepsilon>0$. 则

$$P(|\xi/n-p|\geqslant\varepsilon)\leqslant 2\exp\left(-\frac{n\varepsilon^2}{2\lambda+\varepsilon}\right), \tag{6.16}$$

此处
$$\lambda=\min(p, 1-p).$$

现在来证明 a: 任给$\eta>0$,存在$\varepsilon>0$,使当

$$\begin{aligned}&|u_n(x, x_1, \cdots, x_n)-f_H(x)|<\varepsilon;\\&|u_n(x+1, x_1, \cdots, x_n)-f_H(x+1)|<\varepsilon\end{aligned} \tag{6.17}$$

同时成立时,有(注意x, θ是固定的)

$$|[\delta_n(x|x_1, \cdots, x_n)-\theta]^2-[\delta_H(x)-\theta]^2|<\eta.$$

另一方面,总有

$$[\delta_n(x|x_1, \cdots, x_n)-\theta]^2\leqslant 2(x+1)^2[(n+1)^2+\theta^2],$$

且由引理 6.2$\left(注意到当 0\leqslant Y\leqslant n 时, \left|\frac{Y+1}{n+1}-\frac{Y}{n}\right|\leqslant\frac{1}{n}<\frac{\varepsilon}{2}\right)$,当$n$充分大时,有

$$\begin{aligned}&P(|u_n(x, X_1, \cdots, X_n)-f_H(x)|\geqslant\varepsilon)\\&\quad+P(|u_n(x+1, X_1, \cdots, X_n)-f_H(x+1)|\geqslant\varepsilon)\\&\leqslant 4\exp\left(-\frac{n\varepsilon^2}{8}\right).\end{aligned}$$

于是由以上诸式得到

$$\begin{aligned}&\left[1-4\exp\left(-\frac{n\varepsilon^2}{8}\right)\right]((\delta_H(x)-\theta)^2-\eta)\\&\leqslant E_{(x_1,\cdots,x_n)}[\delta_n(x|X_1, \cdots, X_n)-\theta]^2\\&\leqslant(\delta_H(x)-\theta)^2+\eta+8\exp\left(-\frac{n\varepsilon^2}{8}\right)\\&\quad\cdot[(n+1)^2+\theta^2](x+1)^2.\end{aligned}$$

先令$n\to\infty$,再令$\eta\to 0$,即得

$$\lim_{n\to\infty}E_{(x_1,\cdots,x_n)}[\delta_n(x|X_1, \cdots, X_n)-\theta]^2=(\delta_H(x)-\theta)^2.$$

这就证明了 a.

为证 b, c, 利用引理 6.1, 以 ξ_1, ξ_2, ξ_3 分别记 $X_1, \cdots, X_n$ 中等于 x, $x+1$ 以及既非 x 又非 $x+1$ 的个数, 则 (ξ_1, ξ_2, ξ_3) 服从多项分布 $M(n, p_1, p_2, p_3)$, 其中 $p_1=f_H(x)$, $p_2=f_H(x+1)$. 因为

$$[\delta_n(x|X_1, \cdots, X_n)-\theta]^2 \leqslant 2(x+1)^2\left(\frac{1+\xi_2}{1+\xi_1}\right)^2+2\theta^2,$$

于是按引理 6.1, 得

$$E_{(x_1,\cdots,x_n)}[\delta_n(x|X_1, \cdots, X_n)-\theta]^2$$
$$\leqslant 8(x+1)^2\left[4\left(\frac{f_H(x+1)}{f_H(x)}\right)^2+1\right]+2\theta^2 \triangleq G(x, \theta).$$

显然, $G(x, \theta)$ 与 n 无关, 故 b 已满足. 为证 c, 只需验证

$$E_P\left[(x+1)^2\left(\frac{f_H(x+1)}{f_H(x)}\right)^2\right]<\infty; \tag{6.18}$$

$$E_P(x^2)<\infty; \tag{6.19}$$

$$E_H(\theta^2)<\infty. \tag{6.20}$$

E_P 和 E_H 分别表示在求期望时 X 的分布为 P(见(6.8)), 而 θ 的分布为 H. 由 $\mathscr{H}$ 的定义知(6.20)成立. 又因为

$$E(X^2|\theta)=\theta+\theta^2,$$

故知当 $H\in\mathscr{H}$ 时,

$$E_P(X^2)=E_H(\theta)+E_H(\theta^2)<\infty.$$

剩下证明(6.18). 注意由 Cauchy-Schwarz 不等式, 有

$$(f_H(x+1))^2=\left(\frac{1}{(x+1)!}\right)^2\left[\int_0^\infty e^{-\theta/2}\theta^{x/2}\cdot e^{-\theta/2}\theta^{x/2+1}dH(\theta)\right]^2$$
$$\leqslant\left(\frac{1}{(x+1)!}\right)^2\int_0^\infty e^{-\theta}\theta^x dH(\theta)\int_0^\infty e^{-\theta}\theta^{x+2}dH(\theta)$$
$$=\frac{1}{(x+1)^2}f_H(x)\int_0^\infty e^{-\theta}\frac{\theta^{x+2}}{x!}dH(\theta).$$

于是当 $H\in\mathscr{H}$ 时, 得

$$E_P\left[(X+1)^2\left(\frac{f_H(X+1)}{f_H(x)}\right)^2\right]$$
$$\leqslant E_P\left[\frac{1}{f_H(X)}\int_0^\infty e^{-\theta}\frac{\theta^{X+2}}{X!}dH(\theta)\right]$$
$$=\int_0^\infty\sum_{x=0}^\infty\frac{\theta^{x+2}}{x!}e^{-\theta}dH(\theta)=\int_0^\infty\theta^2 dH(\theta)<\infty.$$

这就证明了 c. 因而证明了 δ_n 确为 θ 的 a. o. 经验 Bayes 估计. 在[25]中,对一般的一维离散指数族证明了类似的结果.

注意在本例中，我们不是先从历史样本 $x_1, \cdots, x_n$ 去对先验分布作一个估计，然后再利用这个估计作为先验分布以构造经验 Bayes 估计. 这一般地只在像例 6.1 那样,当 $\mathscr{H}$ 为一个参数族时才行. 相反,在这里我们先算出 H 已知时的 Bayes 估计(6.9),然后利用如下的事实: 估计 (6.9) 只是通过 X 的边缘分布而依赖 H. 一般地,要 a. o. 经验 Bayes 解存在,这个条件必须满足(不难验证例 6.1 满足这个要求). 下面再举一个例子,但不作仔细论证.

例 6.2 设 $X\sim R(0, \theta)$, $\theta>0$. X 的分布函数和密度函数分别为(以下都是 $x>0$)

$$F_\theta(x)=\begin{cases} x/\theta, & 0<x<\theta, \\ 1, & x\geqslant\theta; \end{cases}$$

$$f_\theta(x)=\begin{cases} \theta^{-1}, & 0<x\leqslant\theta, \\ 0, & x>\theta. \end{cases}$$

当 θ 有先验分布 H 时,X 的边缘分布函数和边缘密度函数分别为

$$K(x)=\int_{(0,x)} dH(\theta)+\int_{(x,\infty)} \frac{x}{\theta}\, dH(\theta);$$

$$k(x)=\int_{[x,\infty)} \frac{1}{\theta}\, dH(\theta).$$

易见 $K(x)=H(x)+xk(x)$. 在平方损失 $L(\theta, a)=(\theta-a)^2$ 之下用公式(3.8),易见在先验分布 H 之下,θ 的 Bayes 估计为

$$\delta_H(x)=\int_{[x,\infty)} dH(\theta)\Big/\int_{[x,\infty)} \frac{1}{\theta}\, dH(\theta)=\frac{1-H(x)}{k(x)}$$

$$=[1-K(x)+xk(x)]/k(x)=x+\varphi(x). \qquad (6.21)$$

此处

$$\varphi(x)=\frac{1-K(x)}{k(x)}, \qquad (6.22)$$

由(6.21)与(6.22)知, $\delta_H(x)$ 只通过 X 的边缘分布(即 $K(x)$)依赖于 H. 为得到 θ 的一个经验 Bayes 估计，只需从历史样本 x_1,

$\cdots$, x_n 估计 $K(x)$ 及 $k(x)$. 一个可用的估计是以

$$K_n(x)=\{x_1,\ \cdots,\ x_n \text{ 中} \leqslant x \text{ 的个数}\}/n$$

估计 $K(x)$. 要估计 $k(x)$, 取 $h_n>0$, 用

$$k_n(x)=[K_n(x)-K_n(x-h_n)]/h_n.$$

为了得到较好的收敛性质, 不能直接以 K_n 和 k_n 代(6.22)中的 K 和 k, 而要作一些截断. 具体说, 选择适当的 $a_n(x)$, 令

$$\varphi_n(x)=\left[\min\left(a_n(x),\ \frac{1-K_n(x)}{k_n(x)}\right)\right]I_{[h_n,\infty)}(x). \tag{6.23}$$

若令
$$\|a_n\|_p=\left[\int_0^\infty |a_n(x)|^p dx\right]^{1/p},\ 1\leqslant p<\infty,$$

$$\|a_n\|_\infty=\inf\{\sup_{x\in N}|a_n(x)|: N \text{ 为任意 } L\text{-零测集}\},$$

则 Fox 证明了: 若

1° 对一切 $x>0$, 有 $\lim\limits_{n\to\infty} a_n(x)=\infty$;

2° $\|a_n\|_1=0(\sqrt{n})$, $\|a_n\|_2^2=0(\sqrt{n}h_n)$, $\|a_n\|_\infty=0(n^{1/4})$;

3° $\lim\limits_{n\to\infty} h_n=0$.

则
$$\delta_n(x|x_1,\ \cdots,\ x_n)=x+\varphi_n(x)$$

为 θ 的 a. o. 经验 Bayes 估计(相对于全体先验分布构成的族 $\mathscr{H}$).

在此不给出这个结果的仔细证明, 读者可参看[21], 在 Fox 的这个工作中, 还有另一些经验 Bayes 估计的例子. 从这些例子以及我们仔细讨论过的 Poisson 分布的例子可看出: 虽然构造一个可能的经验 Bayes 估计不见得很困难, 但要严格证明其为 a. o. 经验 Bayes 估计, 却无一定的方法可循, 并且不容易.

在 $\delta_H(x)$ 不只是通过 X 的边缘分布而依赖于 H 的情况下, a. o. 经验 Bayes 估计多不存在. 一个简单的例子如下: 设 X 服从二项分布 $B(m,\ \theta)$ $(0\leqslant\theta\leqslant1)$, 在平方损失之下, 当先验分布为 H 时, 依公式(3.8)得到 θ 的 Bayes 估计为

$$\delta_H(x)=\int_0^1\binom{m}{x}\theta^{x+1}(1-\theta)^{m-x}dH(\theta)\div$$

$$\int_0^\infty \binom{m}{x} \theta^x (1-\theta)^{m-x} dH(\theta)$$

$$= \frac{x+1}{m+1} f_H(x+1,\ m+1) / f_H(x,\ m).$$

此处 $$f_H(i,\ k) = \int_0^\infty \binom{k}{i} \theta^i (1-\theta)^{k-i} dH(\theta).$$

由于$f_H(x+1,\ m+1)$不是$\{f_H(x,\ m): x=0, 1,\ \cdots,\ m\}$的函数(这一点从直观上看很明显，也可以举例证明). 可以证明: 若不对先验分布族 $\mathscr{H}$ 作一定的限制，θ 的 a. o. 经验 Bayes 估计就不存在，当对 $\mathscr{H}$ 作了一定的限制，例如 $\mathscr{H}$ 为例 4.1 中的 β-分布族时，θ 的 a. o. 经验 Bayes 估计便存在.

三、经验 Bayes 估计的收敛速度

即使 $\{\delta_n = \delta_n(x | x_1,\ \cdots,\ x_n)\}$ 为 a. o. 经验 Bayes 估计，即 (6.3)式成立，其收敛速度也可能很慢，因此从理论和实用的角度看，研究(6.3)式的收敛速度是一项有意义的工作. 目前在这方面，只是在单参指数族的情况，有了一些结果. 主要是 Lin, Singh 和赵林城所作的. 由于牵涉不少枝节且过于专门，在此不拟深入讨论这个问题了. 有兴趣的读者可阅读有关的文献[21]～[24]. 以下不加证明地介绍赵林城所得到的一个结果.

设损失函数为 $L(\theta,\ a) = (\varphi(\theta) - a)^2$，先验分布为 H，且设样本空间为 $\mathscr{X} = \{0,\ 1,\ 2,\ \cdots\}$，$f_\theta(x)$为当参数为 θ 时 $X = x$ 的概率，而 $f(x)$ 为 X 的边缘分布，即 $f(x) = \int_\Theta f_\theta(x) dH(\theta)$. 设 $f(x) > 0$ $(x = 0,\ 1,\ 2,\ \cdots)$. 假定 $\varphi(\theta)$ 的 Bayes 估计 $\delta_H(x)$有形式

$$\delta_H(x) = \sum_{k=0}^{\infty} a_k(x) f(x+k) / f(x). \qquad (6.24)$$

这里 $a_k(x)$ $(k = 0,\ 1,\ 2,\ \cdots)$都是 x 的已知函数. 不难知道，前面讨论过的 Poisson 分布是 (6.24) 的特例，其中 $a_1(x) = x+1$，$a_k(x) \equiv 0$ 当 $k \neq 1$.

定理6.1 假定

1° $A(x) \triangleq \sum_{k=0}^{\infty} a_k^2(x) f(x+k) < \infty$ $(x=0, 1, 2, \cdots)$;

2° $\int_{\Theta} |\varphi(\theta)|^r dH(\theta) < \infty$ (对某个 $r>2$);

3° $\sum_{k=0}^{\infty} \left[1+\frac{A^\lambda(k)}{f^\lambda(k)}\right] f^{1-\lambda}(x) < \infty$ (对某个 $\lambda \in (0, 1)$).

设有历史样本 $x_1, \cdots, x_n$, 用

$$f_n(x) = \frac{1}{n}\{x_1, \cdots, x_n \text{ 中等于 } x \text{ 的个数}\}$$

估计 $f(x)$. 令

$$g_n(x) = \sum_{k=0}^{\infty} a_k(x) f_n(x+k).$$

而定义$\left(\text{约定 } \frac{0}{0}=0;\ \text{当 } c \neq 0 \text{ 时}, \left|\frac{c}{0}\right| = \infty.\right)$

$$\delta_n(x|x_1, \cdots, x_n) = \begin{cases} g_n(x)/f_n(x), \text{当} |g_n(x)/f_n(x)| \leqslant n^\nu; \\ 0, \text{当} |g_n(x)/f_n(x)| > n^\nu. \end{cases}$$

则可以选定适当的 $\nu \in (0, 1)$, 使

$$|R_H^*(\delta_n) - R_H(\delta_H)| = O(n^{-(r-2)\lambda/r}). \tag{6.25}$$

特别是, 若对任何 $r>2$, 条件 2° 成立; 而对任意接近于 1(但小于 1) 的 λ, 条件 3° 成立, 则由(6.25)知, $R_H^*(\delta_n) - R_H(\delta_H)$ 收敛于 0 的速度可任意接近于 $O\left(\frac{1}{n}\right)$.

这个定理的证明及其应用的例子, 可参看赵林城[26].

四、复合判决问题

设某厂生产一种产品, 其质量指标服从分布 $N(\theta, 1)$. 但 θ 并非固定而是随机的. 例如, 可设想为 θ 的值逐日有些波动, 因此可以认为 θ 有一定的概率分布 H. 现在对某特定的 n 天中, 每天抽了一定的样本, 第 i 天为 x_i, 一共得到样本 $x_1, \cdots, x_n$. 要估计这特定的 n 天的 θ 值 $\theta_1, \cdots, \theta_n$.

这个问题看起来与经验 Bayes 的问题提法很相似, 实际上则

不同．差异之处在于这里推断的对象为 $\theta_1, \cdots, \theta_n$，而不是 θ 的当前值 θ_0．因此，本问题的损失函数一般有形状

$$L(\theta_1, \cdots, \theta_n; a_1, \cdots, a_n)=\sum_{i=1}^{n} L_i(\theta_i, a_i).$$

整个问题可以看作是由 n 个“分问题”合并而成的，故得到“复合判决问题”(compound decision problem)的名称．自然，人们立即会合理地提出这样一个问题：把许多分问题合并在一起去考虑，是否有什么利益．这个问题的仔细分析牵涉面较多，在此就不细述了．但有一点要交代一下：在上述提法中，n 个分问题中的 $\theta_1, \cdots, \theta_n$ 是 θ 的 n 个独立“观察”值．因此联合起来以后，利用 $x_1, \cdots, x_n$ 可得到 θ 的先验分布的信息．这样，将分问题合并起来，显然是有利的且可借用经验 Bayes 方法来处理．在一般情况下，n 个分问题中的参数 $\theta_1, \cdots, \theta_n$ 不必有任何关系．在这种情况下，将它们联合起来去处理的好处在何处，就不是很明显的了．有兴趣的读者可参看[26]．

§7 Minimax 估计

作为判决函数的一种优良性准则，Minimax 原则已在第一章提到过了．说起来，Minimax 原则是一个“自成一统”的准则，并不是附属于 Bayes 准则的．特别是，有些 Bayes 学者甚至否认 Minimax 准则是否值得考虑．原因在于，在 Bayes 学派看来，有一些“合理的基本原则”导致 Bayes 准则，而 Minimax 准则与这些“合理的基本原则”有所径庭．当然，也可以举出一些反面的(有利于 Minimax 准则的)论据来．Minimax 准则的研究在数学上的深入程度并不比 Bayes 准则差．何况，即使以“聊备一格”的态度来看待这个问题，Minimax 准则在判决问题中占据一席之地总归是合理的．至于此地将 Minimax 估计作为本章的一节来叙述，是因为这个内容专设一章不大好处理．而本节叙述的求 Minimax 解的方法与 Bayes 方法有密切的联系．与 Minimax 解有关的另

一些内容将在以后有关的章节中叙述.

先回顾一下 Minimax 解的正式定义. 设参数空间为 Θ, 统计判决函数 δ 的风险函数记为 $R(\theta, \delta)$. 注意, 此处并不排除序贯抽样的情况, 因而 $R(\theta, \delta)$ 也可以是指 δ 的包括平均费用在内的总风险. 设 $\mathscr{F}$ 是所考虑的判决函数类(例如, 一切序贯或非序贯的判决函数的类, 一切样本大小为 n 的固定判决函数的类, 或它们的非随机化的判决函数的子类, 等等), 而 $\delta^* \in \mathscr{F}$.

定义 7.1 若

$$\sup\{R(\theta, \delta^*): \theta \in \Theta\} = \inf\{\sup_{\theta \in \Theta} R(\theta, \delta): \delta \in \mathscr{F}\},$$

则 δ^* 称为(所提判决问题)在 $\mathscr{F}$ 中的 Minimax 解.

在序贯的场合, 问题的提法还可以多样化. 例如, 在对平均费用或因"行动错误"而造成的风险中的一个加上一定的约束, 而对另一个使用 Minimax 原则, 等等.

一、用 Bayes 方法求 Minimax 解

定理 7.1 设存在先验分布 H, 使在此先验分布之下的 Bayes 解 δ_H 的风险函数 $R(\theta, \delta_H)$ 在 Θ 上为一个有限常数 c, 则 δ_H 为一 Minimax 解.

证 设 δ_H 不是 Minimax 解, 则将存在判决函数 δ, 使

$$\sup\{R(\theta, \delta): \theta \in \Theta\} < c,$$

即 对一切 $\theta \in \Theta$, 有 $R(\theta, \delta) < R(\theta, \delta_H)$.

这时有

$$R_H(\delta) = \int_\Theta R(\theta, \delta)\, dH(\theta) < \int_\Theta R(\theta, \delta_H)\, dH(\theta) = R_H(\delta_H).$$

因而与 δ_H 为 Bayes 解有矛盾. 定理证毕.

定理 7.2 设 $\{H_n\}$ 为一串先验分布, 其相应的 Bayes 解的序列为 $\{\delta_n\}$, 而判决函数 δ 满足条件

$$\infty > \sup\{R(\theta, \delta): \theta \in \Theta\} \leqslant \limsup_{n \to \infty} R_{H_n}(\delta_n) \triangleq c. \quad (7.1)$$

则 δ 为一个 Minimax 解.

证 用反证法: 设 δ 不为 Minimax 解, 则存在判决函数 δ^*, 使

$$\sup\{R(\theta,\ \delta^*):\theta\in\Theta\}<\sup\{R(\theta,\ \delta):\theta\in\Theta\}\triangleq c'\leqslant c,$$

且 $c'<\infty$. 因而存在 $\varepsilon>0$, 使

$$对一切\ \theta\in\Theta,\ 有\ R(\theta,\ \delta^*)\leqslant c'-2\varepsilon. \tag{7.2}$$

由 c 的定义, 知存在充分大的 N, 使

$$R_{H_N}(\delta_N)\geqslant c'-\varepsilon. \tag{7.3}$$

于是由(7.2)及(7.3), 得

$$R_{H_N}(\delta^*)=\int_\Theta R(\theta,\ \delta^*)dH(\theta)\leqslant c'-2\varepsilon<c'-\varepsilon\leqslant R_{H_N}(\delta_N).$$

这与 δ_N 为先验分布 H_N 之下的 Bayes 解相矛盾. 定理证毕.

定理 7.1 的应用面较窄, 因为找到适合定理条件的先验分布不容易或不可能. 定理 7.2 的应用较广. 用这个方法, 关键在于找到一串合适的先验分布 $\{H_N\}$[1). 在[28]和[29]中也使用这个方法求 Minimax 解. 下面举几个较简单的例子.

例 7.1 变量 X 服从二项分布 $B(n,\ \theta)$, $0\leqslant\theta\leqslant1$. 损失函数为 $L(\theta,\ a)=(\theta-a)^2$. 在例 3.1 中, 我们已找到先验分布 $\beta(\sqrt{n}/2,\ \sqrt{n}/2)$, 其 Bayes 解 (3.19) 的风险函数为常数 (3.20). 故(3.19)就是 θ 在上述损失函数下的 Minimax 估计.

通常用的估计量为 $\delta^*(x)=x/n$, 其风险函数为 $\theta(1-\theta)/n$, 有

$$\sup_{0<\theta<1}R(\theta,\ \delta^*)=\frac{1}{4n}>\frac{n}{4(n+\sqrt{n})^2}.$$

故 $\delta^*(x)=x/n$ 不是 θ 的 Minimax 估计. 但不难验证, 若损失函数改为

$$L(\theta,\ a)=(\theta-a)^2/(\theta(1-\theta)),\ 0<\theta<1,$$

则 x/n 为 θ 的 Minimax 估计.

例 7.2 设 $x_1,\ \cdots,\ x_n$ 为取自 $N(\theta,\ 1)$ 的 iid. 样本, 损失函数为 $L(\theta,\ a)=(\theta-a)^2$. 求 Minimax 解.

取一串先验分布 $N(0,\ \tau^2)$ $(\tau=1,\ 2,\ \cdots)$, 根据例 3.2 知: 在此先验分布下, θ 的 Bayes 估计为(3.21), 其风险函数为 (3.22),

---

1) 在成平的工作[27]中有一些这方面的较复杂的例子, 可供读者参考.

Bayes 风险为 (3.23)，后者当 $\tau\to\infty$ 时有极限 $\frac{1}{n}$，而这正是 $\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$ 的风险，于是(7.1)适合．由定理 7.2 即知 $\bar{X}$ 就是 θ 的 Minimax 估计．

除了这个使用先验分布的方法外，求 Minimax 估计的另一种方法是基于同变估计的概念，这在本书中将有讨论．但这个方法也只适用于某些性质比较特殊的问题．因此，能求出 Minimax 估计的情况为数很少．这也是 Minimax 原则在使用上的一个缺点，与 Bayes 原则相比较：求 Bayes 估计有一般的、容易实施的方法．

二、Minimax 估计为随机化估计的例子

在第一章中介绍随机化判决函数时，我们曾指出过这个概念在应用上意义不大，但在理论上却有一席地位．在假设检验的 Neyman–Pearson 理论(参看[32])中，必须考虑随机化的检验．在 Bayes 方法中随机化判决函数没有作用，但是在 Minimax 估计中则不然．

如果假定损失函数为凸的，则根据 1.6 节中给出的关于非随机化的充分性原则，不难知道，在寻求 Minimax 估计时，可以局限于只考虑非随机化的估计．事实上，这只需注意样本 x 本身是一个充分统计量就够了．

但是，在损失函数非凸时，Minimax 估计可以是随机化的，下面这个有趣的例子是 Hodges 和 Lehmann 的工作．参看[31]．

例 7.3 设 $X\sim B(n,\theta)$，$0\leqslant\theta\leqslant1$，损失函数为 $L(\theta,a)=|\theta-a|^r$，$0<r<1$，r 已知．则 θ 在一切估计类 $\mathscr{F}$ (包括随机化的和非随机化的估计)中的 Minimax 估计必为随机化的

以 $\mathscr{F}_1$ 记 θ 的一切非随机化估计．我们先证明：θ 的在 $\mathscr{F}$ 中的 Minimax 估计和 θ 在 $\mathscr{F}_1$ 中的 Minimax 估计都存在　先考虑 $\mathscr{F}$ 的问题．因为样本空间只包含 $n+1$ 个点 $0,1,\cdots,n$．任一随机化估计可表为

$$F=(F_0,\ F_1,\ \cdots,\ F_n)$$

的形状. 这里对每个 $i(i=0, 1, \cdots, n)$, F_i 是 [0, 1] 区间上的一个概率分布. 每当抽样得到 $x=i$, 就根据分布 F_i 从[0, 1] 中抽出一个点作为 θ 的估计. 易见

$$R(\theta, F)=\sum_{i=0}^{n} b(n, i, \theta)\int_0^1|\theta-y|^r dF_i(y), \tag{7.4}$$

此处

$$b(n, i, \theta)=\binom{n}{i}\theta^i(1-\theta)^{n-i}.$$

显然, 对任何 F 及 $\theta\in[0, 1]$, 有 $R(\theta, F)\leqslant 1$. 故

$$m(F)=\sup_{0<\theta<1} R(\theta, F)\leqslant 1,$$

而

$$m=\inf_F m(F)\leqslant 1. \tag{7.5}$$

这里 inf 对一切可能的 F 取. 找一串 $\{F^{(k)}=(F_0^{(k)}, \cdots, F_n^{(k)}), k=1, 2, \cdots\}$, 使

$$\lim_{k\to\infty} m(F^{(k)})=m. \tag{7.6}$$

固定 i, 由于 $\{F_i^{(k)}, k=1, 2, \cdots\}$ 为一串概率分布, 其概率全集中在[0, 1]内, 故依 Helly 定理, 从其中可取出一个子序列, 依分布收敛(记为 $\xrightarrow{c}$)于某个分布 F_i^*, 显然 F_i^* 的概率全集中在[0, 1]内. 由于 i 只取 $n+1$ 个值 0, 1, ⋯, n, 可找到自然数的一个子序列, 不妨假定就是 {1, 2, ⋯} 本身, 以及其概率全集中在[0, 1]的分布 $F_i^*(i=0, 1, \cdots, n)$, 使

$$F_i^{(k)}\to F_i^* \quad (\text{当 } k\to\infty;\ i=0, 1, \cdots, n). \tag{7.7}$$

现考虑随机化的估计

$$F^*=(F_0^*, F_1^*, \cdots, F_n^*), \tag{7.8}$$

则有

$$\begin{aligned}R(\theta, F^*)&=\sum_{i=0}^{n} b(n, i, \theta)\int_0^1|\theta-y|^r dF_i^*(y)\\&=\lim_{k\to\infty}\sum_{i=0}^{n} b(n, i, \theta)\int_0^1|\theta-y|^r dF_i^{(k)}(y)\\&=\lim_{k\to\infty} R(\theta, F^{(k)})\leqslant\lim_{k\to\infty} m(F^{(k)})=m,\end{aligned}$$

对 $\theta\in[0, 1]$.

因而 $m(F^*)=\sup\limits_{0\leqslant\theta\leqslant1}R(\theta,\ F^*)\leqslant m$. 这便证明了 F^* 是 θ 的随机化 Minimax 估计.

$\mathscr{F}_1$ 中 Minimax 估计存在的证明更简单. 记

$$\widetilde{m}=\inf\{m(\delta):\delta\in\mathscr{F}_1\}. \tag{7.9}$$

此处仍如前, $m(\delta)=\sup\limits_{0\leqslant\theta\leqslant1}R(\theta,\ \delta)$. 找一串 $\{\delta_k:\ k=1,\ 2,\ \cdots\}$, 使 $\lim\limits_{k\to\infty}m(\delta_k)=\widetilde{m}$. 由于 $\delta_k(i)$ 当 $i=0,\ 1,\ \cdots,\ n$ 及 $k=1,\ 2,\ \cdots$ 全在 $[0,\ 1]$ 内, 存在自然数的子序列, 不妨设就是 $\{1,\ 2,\ \cdots\}$ 本身, 使 $\lim\limits_{k\to\infty}\delta_k(i)$ 存在且记为 $d_i(i=0,\ 1,\cdots,\ n)$. 引进估计 $\delta^*:\delta^*(i)=d_i(i=0,\ 1,\ \cdots,\ n)$, 则不难证明 $m(\delta^*)=\widetilde{m}$, 即 δ^* 为 $\mathscr{F}_1$ 中的 Minimax 估计.

现证(7.9)中的 $\widetilde{m}$ 小于 1. 事实上, 对任何 $\delta\in\mathscr{F}_1$, 有

$$R(\theta,\ \delta)=\sum_{i=0}^{n}b(n,\ i,\ \theta)\,|\theta-\delta(i)|^r.$$

取 δ, 使 $\delta(0)=1-\delta(n)=0$, 当 $i=1,\ \cdots,\ n-1$ 时, $0\leqslant\delta(i)\leqslant1$. 则有 $R(0,\ \delta)=R(1,\ \delta)=0$. 又因 $0\leqslant\delta(i)\leqslant1$, 有

$$\begin{aligned}R(\theta,\ \delta)&\leqslant\max\{|\theta|^r,\ |1-\theta|^r\}\sum_{i=0}^{n}b(n,\ i,\ \theta)\\&<\sum_{i=0}^{n}b(n,\ i,\ \theta)=1,\ \text{当 } 0<\theta<1.\end{aligned}$$

再由 $R(\theta,\ \delta)$ 作为 θ 的函数在 $[0,\ 1]$ 上的连续性, 有

$$\sup_{0\leqslant\theta\leqslant1}R(\theta,\ \delta)<1,$$

故更有 $\widetilde{m}<1$.

以下来证 $\widetilde{m}>m$. 证明了这一点, 也就完成了本题的讨论.

由 $m(\delta^*)=\widetilde{m}$ 知 $\delta^*(0)<1$. 因若 $\delta^*(0)=1$, 将有

$$R(0,\ \delta)=|0-\delta^*(0)|^r=1.$$

而 $m(\delta^*)\geqslant1$, 这与已证的 $\widetilde{m}<1$ 不合. 现有

$$\begin{aligned}R(\theta,\ \delta^*)=&\textstyle\sum^{(1)}b(n,\ i,\ \theta)\,|\theta-\delta^*(0)|^r\\&+\textstyle\sum^{(2)}b(n,\ i,\ \theta)\,|\theta-\delta^*(i)|^r\triangleq I_1+I_2.\end{aligned}$$

此处 $\sum^{(1)}$ 和 $\sum^{(2)}$ 分别表示对满足 $\delta^*(i)=\delta^*(0)$ 的 i 求和及对满足 $\delta^*(i)\neq\delta^*(0)$ 的 i 求和. 记 $R'(\theta,\ \delta^*)=dR/d\theta$. 在 $\delta^*(0)$ 的邻

域内，显然 $dI_2/d\theta$ 有界. 注意到 $\delta^*(0)<1$，在 $\theta>\delta^*(0)$ 而 $\theta-\delta^*(0)$ 充分小时，有 $b(n, 0, \theta)>0$. 于是当 $\theta>\delta^*(0)$ 且与 $\delta^*(0)$ 很接近时，$dI_1/d\theta$ 接近 ∞. 这说明了：在 $\delta^*(0)$ 右边很小的范围内，$R(\theta, \delta^*)$ 为 θ 的严增函数. 由于 $\sup\limits_{0\leq\theta\leq 1} R(\theta, \delta^*)=\widetilde{m}$，有 $R(\delta^*(0), \delta)<\widetilde{m}$. 类似的推理可证明 $R(\delta^*(n), \delta)<\widetilde{m}$. 由于 $R(\theta, \delta^*)$ 为 θ 的连续函数，存在 $\varepsilon>0$, $\beta>0$，使

$$\text{当 } |\theta-\delta^*(0)|<\beta, \text{或} |\theta-\delta^*(n)|<\beta \text{ 时,}$$
$$R(\theta, \delta^*)<\widetilde{m}-2\varepsilon. \tag{7.10}$$

现定义随机化的估计 δ 如下：取定 $\alpha\in(0, \beta)$，而令

$$\delta(\delta^*(x)|x)=1, \text{ 当 } x=1, \cdots, n-1 \text{ 时;}$$
$$\delta(\delta^*(0)+\alpha|0)=\delta(\delta^*(0)-\alpha|0)=\frac{1}{2};$$
$$\delta(\delta^*(n)+\alpha|n)=\delta(\delta^*(n)-\alpha|n)=\frac{1}{2}.$$

则有

$$\begin{aligned}
&R(\theta, \delta)-R(\theta, \delta^*)\\
&\quad=(1-\theta)^n\left[\frac{1}{2}|\theta-\delta^*(0)+\alpha|^r\right.\\
&\qquad\left.+\frac{1}{2}|\theta-\delta^*(0)-\alpha|^r-|\theta-\delta^*(0)|^r\right]\\
&\qquad+\theta^n\left[\frac{1}{2}|\theta-\delta^*(n)+\alpha|^r\right.\\
&\qquad\left.+\frac{1}{2}|\theta-\delta^*(n)-\alpha|^r-|\theta-\delta^*(n)|^r\right]\\
&\quad=J_1+J_2.
\end{aligned} \tag{7.11}$$

取 $\alpha\in(0, \beta)$ 充分小，显然可使

$$|J_1|+|J_2|<\varepsilon, \text{对一切 } \theta\in[0, 1].$$

因此，由(7.10)，得

$$\text{当 } |\theta-\delta^*(0)|\leq\alpha \text{ 或 } |\theta-\delta^*(n)|\leq\alpha \text{ 时,}$$
$$R(\theta, \delta)\leq R(\theta, \delta^*)+\varepsilon<\widetilde{m}-\varepsilon. \tag{7.12}$$

在 $|\theta-\delta^*(0)|>\alpha$ 和 $|\theta-\delta^*(n)|>\alpha$ 同时成立时，则因函数 $-|x|^r$ 在 $x\geq 0$ 处为严凸(注意 $0<r<1$)，立即得出 J_1 与 J_2 的表

达式中的方括号内的量都小于0. 由于θ和$1-\theta$不同时为0，当$|\theta-\delta^*(0)|>\alpha$和$|\theta-\delta^*(n)|>\alpha$都成立时，有

$$R(\theta,\ \delta)<R(\theta,\ \delta^*).$$

由此及(7.12)，并注意到$R(\theta,\ \delta^*)$和$R(\theta,\ \delta)$都是θ的连续函数，则得

$$m(\delta)=\sup_{0\leqslant\theta\leqslant1}R(\theta,\ \delta)<\sup_{0<\theta<1}R(\theta,\ \delta^*)=\widetilde{m}.$$

故更有$m=\inf_{\delta}m(\delta)<\widetilde{m}$. 这就证明了：$\mathscr{F}$中的Minimax估计决不能落在$\mathscr{F}_1$内.

问题与习题

1. 为了定义Bayes风险$R_B(\delta)$，需要证明$R(\theta,\ \delta)$的$\mathscr{B}_\Theta$-可测性(此处δ可以是随机化的). 问：为证此命题需对损失函数$L(\theta,\ a)$加上什么条件？如何证明？

2. 证明由(3.8)式定义的平方损失下的Bayes解为$\mathscr{B}_x$-可测的.

3. 证明引理3.2，并由它推出Bayes解.

4. 证明4.3(一)末尾处的结论；仔细证明例4.4～例4.6中未证的结论.

5. a) 设$X\sim B(n,\ \theta)$，$0\leqslant\theta\leqslant1$，损失函数为$L(\theta,\ a)=(\theta-a)^2$. 证明$\delta(x)=x/n$为Bayes估计，并指出相应的先验分布；若$\theta$限制为$0<\theta<1$，则$\delta(x)$不是Bayes估计，但它在两种意义下都是广义Bayes估计.

b) 设$X\sim\mathscr{P}(\theta)$，$\mathscr{P}(\theta)$为参数为θ的Poisson分布，$\theta\geqslant0$. 试证明：$\delta(x)=x$为在平方损失下、在某个先验分布下θ的Bayes估计. 但若θ限制为$\theta>0$，则$\delta(x)$不是Bayes估计，不过在两种意义下都是广义Bayes估计.

6. 若$\delta(x)$为θ的无偏估计，且对任何$\theta\in\Theta$，有$0<\mathrm{Var}_\theta(x)<c$，$c<\infty$为常数. 试证明：$\delta(x)$决非$\theta$的Bayes估计；利用本题的结果考察第5题及例5.1.

7. 设$X\sim B(n,\ \theta_1)$，$Y\sim B(n,\ \theta_2)$. X，Y独立. $(\theta_1,\ \theta_2)$的先验分布为：θ_1、θ_2独立，都服从均匀分布$R(0,\ 1)$. 损失函数为$L(\theta_1,\ \theta_2,\ a)=[(\theta_2-\theta_1)-a]^2$(这表示要估计$\theta_2-\theta_1$). 求Bayes估计.

8. 举一个简单例子证明：不同的先验分布可以导致完全相同的Bayes解.

9．设 $X \sim B(n, \theta)$，已证在平方损失 $L(\theta, a)=(\theta-a)^2$ 之下，

$$\delta(x)=\frac{x}{n+\sqrt{n}}+\frac{\sqrt{n}}{2(n+\sqrt{n})}$$

为 θ 的 Minimax 估计．利用这个事实证明 $\delta(x)$ 为 θ 的可容许估计．

10．设 $X \sim B(n, \theta)$．利用 x/n 为在损失函数 $L(\theta, a)=(\theta-a)^2$ 之下，适当的广义先验分布的广义 Bayes 解这个事实，证明它是 θ 的可容许估计．问：这个方法可否用于证明 $\bar{x}$ 作为 $N(\theta, 1)$的 θ 的估计(损失函数为$(a-\theta)^2$)的可容许性？

11．设 $x_1, \cdots, x_n$ 为抽自均匀分布 $R(0, \theta)$，$\theta>0$ 的 *iid*.样本，损失函数为$(\theta-a)^2$．求 θ 的 Minimax 估计．

12．设 $x_1, \cdots, x_n$ 为抽自 $R\left(\theta-\frac{1}{2}, \theta+\frac{1}{2}\right)(-\infty<\theta<\infty)$的 *iid*.样本，损失函数为$(a-\theta)^2$．求 θ 的 Minimax 估计．

13．指定实数 μ 和 $\tau^2>0$．以 $\mathscr{F}$ 记一切一维先验分布的族，其均值为 μ，方差为 τ^2．设 $X \sim N(\theta, 1)$．试证明

$$\delta^*(x)=\frac{\tau^2}{1+\tau^2}x+\frac{\mu}{1+\tau^2}$$

有这样的性质：它使 $k(\delta)=\sup\{R_H(\delta): H \in \mathscr{F}\}$ 达到最小(即对任何 δ，有 $k(\delta^*) \leqslant k(\delta)$)．

14．证明：若某先验分布之下的 Bayes 解 δ 唯一，则 δ 必是可容许的．用这个事实证明：若 $X \sim N(\theta, 1)$，则在平方损失之下，$\delta_{a,b}(x)=ax+b$ 当 $0 \leqslant a<1$ 时，必为可容许的．

15．举例说明：相对于广义先验分布的(广义)Bayes 解不必是可容许的$\left[\text{注：考虑 } x \sim G\left(\frac{1}{\theta}, 2\right)\text{，平方损失(}\Theta\text{ 为}(0, \infty)\text{)，广义先验分布 } d\theta/\theta\right]$．

16．设 $X \sim N(\theta, 1)$，先验分布族为 $\theta \sim N(\mu, \tau^2)$，$-\infty<\mu<\infty$，$0<\tau^2<\infty$．试找出一个针对这先验分布族的、$\theta$ 的经验 Bayes 估计（损失为平方$(\theta-a)^2$)．

17．设 $X \sim N(\theta, 1)$，损失为$(\theta-a)^2$．试证明：在任一先验分布 G 之下，θ 的 Bayes 估计可表为

$$\delta_G(x)=x+f_G'(x)/f_G(x).$$

此处 f_G 为 X 的边缘分布密度．

18．设 X 的分布为一维连续指数族

$$f_\theta(x)dx=c(\theta)e^{Q(\theta)t(x)}h(x)dx, \ \theta \in \Theta.$$

此处 $Q(\theta)$ 在 Θ 上是非降的，又 $t(x)$对 x 可微．试证明在平方损失下对任一先验分布 G，θ 的 Bayes 估计 $\delta_G(x)$为 x 的非降函数．

19. 设 $X\sim N(\theta,\ 1)$，先验分布为 $N(0,\ 1)$，损失为

$$L(\theta,\ a)=\exp(3\theta^2/4)(\theta-a)^2.$$

试证明: a) 在定义 2.2 之下，任一估计 $\delta(x)$ 都是 Bayes 估计; b) 但是若按"后验风险最小"的方式来定义 Bayes 估计，则这样的 Bayes 估计 $\delta^*(x)$ 是唯一的. 试求出这个估计; c)证明 δ^* 不可容许(注: 与 $\delta_1(x)=x$ 比较); d)如果将损失函数改为 $\exp(\theta^2/4)(\theta-a)^2$，而其他不变，情况如何?

20. 设 $X\sim P_\theta$, θ 为一维参数，且 $E_\theta(X)=\theta$. 损失函数为$(\theta-a)^2$, θ 的先验分布 G 满足条件

$$\int_{-\infty}^{\infty}\theta^2\,dG(\theta)<\infty.$$

在 θ 固定时作两次独立抽样 $X,\ Y$. 以 $\delta_G(x)$记在样本 X 的基础上所作出的 θ 的 Bayes 估计. 试证明:

$$\delta_G(X)=E^*(Y|X).$$

这里 E^* 表示在求条件期望时，(X, Y)应理解为边缘分布，即其分布 P^* 为

$$P^*(X\in A,\ Y\in B)=\int_\Theta P_\theta(A)P_\theta(B)dG(\theta)\quad (\Theta \text{ 为参数空间}).$$

21. 设参数空间和行动空间都只包含两个点$\{\theta_1,\theta_2\}(-\infty<\theta_1<\theta_2<\infty)$. 当 $\theta=\theta_i$ 时，X 的分布是 $N(\theta_i,\ 1)$假定损失函数是"0-1"的，即 $L(\theta,\ a)=0$ 或 1，视 $\theta=a$ 或否而定. 试对本问题构造出一个(针对一切先验分布的族的)a. o. 经验 Bayes 解(当然，本问题实质上是一个检验问题).

22. 设 $X\sim P_\theta$, θ 为一维参数，损失函数为$(a-\theta)^2$. 先验分布族为

$$\mathscr{F}=\left\{G:\int_\Theta\theta^2\,dG<\infty\right\}\quad (\Theta \text{ 为参数空间}),$$

以 δ_G 记先验分布为 G 时，θ 的 Bayes 估计，而 F_G 为 X (在先验分布 G 之下)的边缘分布. 试证明

$$\text{对任何 } G\in\mathscr{F}\text{，有 }\int_{\mathscr{X}}\delta_G^2(x)dF_G(x)<\infty.$$

现假定

a) $\lim\limits_{L\to\infty}\int_{\{x:|\delta_G(x)|>L\}}\delta_G^2(x)dF_G(x)=0$ 对 $G\in\mathscr{F}$ 一致成立(即: 当 $L\to\infty$ 时，这积分对 $\mathscr{F}$ 中的 G 一致趋于 0);

b) $\{\delta_n(x|x_1,\ \cdots,\ x_n)\}$为一串经验 Bayes 估计，使对每个 x 及 $G\in\mathscr{F}$，有

$$\delta_n(x|X_1,\ \cdots,\ X_n)\xrightarrow{P}\delta_G(x).$$

又对任何 $a>0$ 及 b，以$[b]_a$记 $a,\ b$ 或 $-a$，视 $b>a$, $|b|\leqslant a$ 或$b<-a$而定. 试证明: 可适当选择 L_n，使

$$\delta_n^*(x|x_1,\ \cdots,\ x_n)=[\delta_n(x|x_1,\ \cdots,\ x_n)]_{L_n}$$

为相对于先验分布族 $\mathscr{F}$ 的 a. o. 经验 Bayes 估计.

23. 设 $X\sim N_p(\theta,\Sigma)$, $p\geqslant 3$, $\Sigma>0$ 已知. 取

$$m=(p-2)/2; \qquad \rho=m/(m+1).$$

又选定 p 维向量 μ 及 p 阶正定方阵 A. 给出 θ 的广义先验密度为

$$dG(\theta)=g(\theta)d\theta$$
$$=(2\pi)^{-p/2}\left[\int_0^1 [|B(\lambda)|]^{-1/2}\exp\left\{-\frac{1}{2}(\theta-\mu)'B^{-1}(\lambda)(\theta-\mu)\right\}\lambda^{-2}d\lambda\right]d\theta,$$

此处 $B(\lambda)=\rho[\lambda^{-1}(\Sigma+A)-\Sigma]$, 而 $|B(\lambda)|$ 表示 $B(\lambda)$ 的行列式. 试证明在此广义先验分布下, 依"后验风险"最小原则算出的(广义)Bayes 解为

$$\delta(x)=x-\frac{r((x-\mu)'(\Sigma+A)^{-1}(x-\mu)/\rho)\Sigma(\Sigma+A)^{-1}}{(x-\mu)'(\Sigma+A)^{-1}(x-\mu)}(x-\mu),$$

此处

$$r(v)=2n\left(1-\frac{1}{n\int_0^1\lambda^{n-1}e^{-(\lambda-1)v/2}d\lambda}\right).$$

为了叙述以下两题, 先介绍一个概念: 设 X 的分布函数(给定 θ 时)为 $F(x|\theta)$, 而 θ 的先验分布为 $dG(\theta)$, 则 X 的边缘分布函数为

$$F_G(x)=\int_\Theta F(x|\theta)dG(\theta).$$

若 $\mathscr{F}$ 为 θ 的一个先验分布族, 且当 $G_1\in\mathscr{F}$, $G_2\in\mathscr{F}$, 而 $F_{G_1}\equiv F_{G_2}$ 时, 必有 $G_1=G_2$, 则称 G(局限于 $\mathscr{F}$ 时)是"*可以辨识*"的. 当 $\mathscr{F}$ 为一切先验分布的族时, 就称 G *可以辨识*.

24. a) 设 $F(x|\theta)$ 是参数为 θ 的 Poisson 分布($\theta>0$), 试证明 G 是可以辨识的.

b) 设 $F(x|\theta)$ 为二项分布 $B(N,\theta)$, $0<\theta<1$. 试证明: 若 $N\geqslant 2$, 则对先验分布族 $\mathscr{F}=\{\beta(a,b): a>0, b>0\}$, G 可以辨识. 但对一切先验分布族则不然.

25. 设 $f(x)dx$ 为一维密度函数, 其特征函数处处不为 0. 而

$$F(x|\theta)=\int_{-\infty}^{x-\theta}f(t)dt \quad (-\infty<\theta<\infty).$$

试证明 G 可以辨识.

参 考 文 献

[1] J. O. Berger: Statistical Decision Theory. Springer-Verlag, 1980.

[2] H. Rubin: Axiomatic development of rational behavior, Technical Report., Dept. of Statistics, Purdue Univ., 1974.

[3] T. S. Ferguson: Mathematical Statistics, A Decision-Theoretic Approach, Academic Press. 1967.

[4] L. J. Savage: The Foundations of Statistical Inference,Methuen, 1962.
[5] M. H. DeGroot: Optimal Statistical Decisions, McGraw-Hill, 1970.
[6] D. V. Lindley: Bayesian Statistics, A Review. S. I. A. M., 1971.
[7] S. Zaks: Theory of Statistical Inference, John Wiley, 1971.
[8] D. Blackwell, M. A. Girshick: Theory of Games and Statistical Decisions, John Wiley, 1954.
[9] A. Wald: Statistical Decision Functions, John Wiley, 1950.
[10] R. H. Farrell: Towards a theory of generalized Bayes tests. Ann. Math. Statist., 1968, p. 1.
[11] J. Sacks: Generalized Bayes Solutions in estimation problems, Ann. Math. Statist., 1963, p. 751.
[12] R. H. Farrell: Weak limits of sequences of Bayes procedures in estimation theory, Proc. V. Berkeley Symp. Math. Statist. I, p. 83.
[13] P. R. Halmos: Measure Theory, Van Nostrand, 1950.
[14] H. Robbins: The empirical Bayes approach to Statistics, Proc. III Berkeley Symp. Math. Statist., I. p. 157.
[15] J. S. Maritz: Empirical Bayes Methods, Methuen, 1970.
[16] H. E. Robbins: The empirical Bayes approach to statistical decision problems. Ann. Math. Statist., 1964, p. 1.
[17] J. Neyman: Two breakthroughs in the theory of statistical decision making, Rev. Inst. Internat. Statist., 1962, p. 11.
[18] M. V. Jones: Jr. Contributions to the theory of nonparametric empirical Bayes procedures in statistics, Columbia Univ. Disertation, 1956.
[19] W. Hoeffding: Probability inequalities for sums of bounded random variables, J. Amer. Statist. Assoc. 1963, p. 13.
[20] R. J. Fox: Solutions to empirical Bayes Squared error loss estimation problems, Ann. Statist., 1978, p. 846.
[21] P. E. Lin: Rates of convergence in empirical Bayes estimation problems: Discrete case. Ann. Inst. Statist. Math. 1972, p. 319.
[22] P. E. Lin: Rates of convergence in empirical Bayes estimation problems: Continuous case, Ann. Statist., 1975, p. 155.
[23] R. S. Singh: Empirical Bayes estimation with convergence rates in noncontinuous Lebesgue exponential families, Ann. Statist., 1976, p. 431.
[24] R. S. Singh: Empirical Bayes estimation in Lebesgueexponential families with rates near the best possible rate, Ann. Statist., 1979, p. 890.
[25] 陈希孺: Asymptotically optimal empirical Bayes estimates for parameter of one-dimensional exponential families.
[26] J. B. Copas: Compound Decisions and Empirical Bayes. J. Royal Statist. Soc. Series B 1969 p. 397.

[27] 成平:《指数族分布的参数的极小极大化估计》. 数学学报 1964 年, p. 252.
[28] 陈希孺:《当误差分布已知时回归系数的 Minimax 估计》. 数学进展, 1964, p. 450.
[29] 陈希孺: 《t-区间估计的 Minimax 性质》. 科学通报. 1966 年第 2 期.
[30] E. L. Lehmann: Testing Statistical Hypothesis, John Wiley, 1959.
[31] J. L. Hodges: E. L. Lehmann. Some problems in minimax estimation, Ann. Math. Statist., 1950, p. 182.

第五章　参数估计的大样本理论

参数估计问题是利用对总体的抽样结果来估计总体的真参数或其函数的．当样本的大小(容量)无限增大时，估计是否逼近真值？以什么样的速度逼近？它的极限分布是什么？这显然是数理统计中的一个根本问题，无论是从理论意义上或是从应用角度上来看，它都占有很重要的地位．因此，它一直是数理统计学家所注意研究的对象．历史上的著名人物，如 Gauss, Pearson, Fisher, Neyman, Cramér, Wald 等在大样本理论中都有过杰出的贡献．在近代，它仍然是数理统计学家所注意研究的项目．本章将介绍这方面的基本概念和内容，也就是回答开头所提到的三个问题．一个估计在子样大小无限增大时，能无限逼近真值，则称它为**相合估计**．从直观上说，估计的相合性是在大样本理论中或者说在渐近理论中对估计的起码要求．第二个问题是收敛速度问题，这里首先要分辨出收敛速度的阶或者量级，其次是在同一量级中如何进一步比较优劣，在比较优劣时又可以从各种不同的角度上来考虑，从而产生种种不同的标准．第三个问题是，在同一量级中，估计的极限分布是什么？这个问题又与第二个问题有密切联系，从极限分布的性状，可以比较同一收敛量级的估计的优劣．上述这些估计优劣的比较，不仅要求所比较的收敛的阶或者量级相同，还得限制在某类估计中来比较．比较好的估计一般称之为**渐近有效估计**．当然，由于采用的比较标准不同，会有种种不同的渐近有效估计的定义．下面分节对上述的基本原则和内容进行论述．

§1　概论及预备知识

一、相合估计的定义

我们设 $(\mathscr{X}_n, \mathscr{B}_{x_n}, \mathscr{P}_n)$ $(n=1, 2, \cdots)$ 是一串概率空间，且

$(\mathscr{X}_n, \mathscr{B}_{x_n})\subset(\mathscr{X}_{n+1}, \mathscr{B}_{x_{n+1}})$，其中 $\mathscr{P}_n=\{P_{n,\theta}, \theta\in\Theta\}$ $(n=1, 2, \cdots)$，Θ 为参数空间．在本章中，我们总假定 Θ 是 p 维欧氏空间上的开集或开区域，并假设 $\mathscr{P}_n\ll\mu_n$，μ_n 为 $(\mathscr{X}_n, \mathscr{B}_{x_n})$ 上 σ-有限测度．记 $p_n(x_n, \theta)=dP_{n,\theta}(x_n)/d\mu_n$.

在本书中，讨论的情况绝大多数是 $(\mathscr{X}_n, \mathscr{B}_n)=(\mathscr{X}_1\times\mathscr{X}_1\times\cdots\times\mathscr{X}_1, \mathscr{B}_{x_1}\times\mathscr{B}_{x_1}\times\cdots\times\mathscr{B}_{x_1})$，$P_{n,\theta}=\prod\limits_{i=1}^{n}P_{1,\theta}(x_i)$ $(n=1, 2, \cdots)$. 也就是说，从总体 $(\mathscr{X}_1, \mathscr{B}_{x_1}, P_{1,\theta})$ $(\theta\in\Theta)$ 中独立随机地抽样，其结果为 $X_1, X_2, \cdots$，即为一串 iid. 随机变量．在个别场合，抽样结果不是来自同一总体，而是来自不同总体，例如在线性模型中就是这样，它就是前面提到的一般情况．

设 $\hat{\theta}_n$ 为 $(\mathscr{X}_n, \mathscr{B}_{x_n})\to(R^p, \mathscr{B}^p)$ 可测变换，则称 $\{\hat{\theta}_n\}$ 为 θ 的估计序列，其中 $\hat{\theta}_n$ 仅依赖样本值 $x_n\in\mathscr{X}_n$.

定义 1.1 我们称 $\{\hat{\theta}_n\}$ 为 θ 的相合估计，如果对一切 $\theta\in\Theta$ 以及任一正数 ε，总有

$$\lim_{n\to\infty}P_{n,\theta}\{\|\hat{\theta}_n-\theta\|\geqslant\varepsilon\}=0. \tag{1.1}$$

此处 $\|\hat{\theta}_n-\theta\|$ 表示 $\hat{\theta}_n$ 与 θ 的欧氏距离．(1.1) 式也记为 $\hat{\theta}_n\xrightarrow{P}\theta$.

如果进一步假设对一切 $\theta\in\Theta$，总有

$$\lim_{n\to\infty}P_\theta\left(\bigcup_{k=n}^{\infty}\{\|\hat{\theta}_k-\theta\|\geqslant\varepsilon\}\right)=0. \tag{1.2}$$

其中 P_θ 是指在 $(\mathscr{X}_\infty, \mathscr{B}_{x_\infty})$ 上的由 $\{P_{n,\theta}, n=1, 2, \cdots\}$ 产生的概率测度，则称 $\hat{\theta}_n$ 为 θ 的强相合估计，记为 $\hat{\theta}_n\to\theta$ a. s. P_θ.

不难看出，若 $\hat{\theta}_n$ 为 θ 的强相合估计，则必为在(1.1)式意义下的相合估计，因此，我们又称这种相合估计为弱相合估计．

定义 1.2 我们称 $\{\hat{\theta}_n\}$ 均方收敛于 θ，如果对一切 $\theta\in\Theta$，有

$$\lim_{n\to\infty}E_\theta\{\|\hat{\theta}_n-\theta\|^2\}=0. \tag{1.3}$$

不难验证：若 $\{\hat{\theta}_n\}$ 均方收敛于 θ，则 $\{\hat{\theta}_n\}$ 是 θ 的相合估计．这是因为对任意的 $\varepsilon>0$，有

$$P_\theta\{\|\hat{\theta}_n-\theta\|\geqslant\varepsilon\}\leqslant\frac{1}{\varepsilon^2}E_\theta\{\|\hat{\theta}_n-\theta\|^2\}\to 0\quad(n\to\infty).$$

定义 1.3 设 $\{\xi_n\}$ 为一随机变量序列，如果存在 $\{c_n>0\}$，且 $\lim\limits_{n\to\infty}c_n=\infty$，使对任意的 $\varepsilon>0$，有

$$\lim_{n\to\infty}P\{|c_n\xi_n|\geqslant\varepsilon\}=0. \tag{1.4}$$

则说 $\{\xi_n\}$ 比 c_n^{-1} 以更快的速度依概率趋于零，记为 $\xi_n=o_P(c_n^{-1})$.

如果仅使得

$$\lim_{L\to\infty}\limsup_{n\to\infty}P\{|c_n\xi_n|\geqslant L\}=0, \tag{1.5}$$

则称 $\{\xi_n\}$ 以不低于 c_n^{-1} 的速度依概率地趋于零，记为

$$\xi_n=O_P(c_n^{-1}).$$

定义 1.4 我们称 $\{\hat{\theta}_n\}$ 为 θ 的 $\{c_n^{-1}\}$ 阶相合估计，如果对任一 $\theta\in\Theta$ 及任一 $\varepsilon>0$，存在充分大的 L_0，使当 $L\geqslant L_0$ 时，有

$$\limsup_{n\to\infty}P_{n,\theta}\{\|c_n(\hat{\theta}_n-\theta)\|\geqslant L\}<\varepsilon, \tag{1.6}$$

此处 c_n 为一串非奇异阵，当 $n\to\infty$ 时，$c_n^{-1}\to 0$.

读者不难发现，这些定义也可以拓广到强相合估计上去.由于本书较少用到，在此就不叙述了.

定义 1.4 是比较 $\hat{\theta}_n$ 收敛于 θ 的速度. 另一方面，我们也可以比较 $P_\theta\{\|\hat{\theta}_n-\theta\|>\varepsilon\}$ 这个概率收敛于 0 的速度.

定义 1.5 如果对任意小的 $\varepsilon>0$ 及任意的 $\theta\in\Theta$，当 $\{c_n>0\}$ 时，$\lim\limits_{n\to\infty}c_n=\infty$，而

$$\limsup_{n\to\infty}c_nP_{n,\theta}\{\|\hat{\theta}_n-\theta\|\geqslant\varepsilon\}<\infty. \tag{1.7}$$

则说相合估计 $\hat{\theta}_n$ 收敛于真值 θ 的概率是以 $\{c_n^{-1}\}$ 的速度趋于 0，或简称为 c_n^{-1} 速度.

下面，我们着重研究它的指数速度. 因为不少常见的估计是以指数速度收敛的.

为了确切给出渐近有效估计的定义，还必须有若干预备知识. 从直观上想，一个估计的好坏，主要看这个估计能否更准确地区别参数的真值和非真值. 这就涉及区分真值与非真值的标志以及在

已得到样本时求找精确区分的办法. 前者将引出广义密度距离的概念, 随之而产生 Kullback-Leibler 信息函数, 第二个问题就涉及到假设检验, 特别是 Neyman-Pearson 基本引理. 由于本章主要涉及随机序列的收敛问题, 也有必要回顾一下这方面的某些结果. 下面就这三个方面加以讨论, 然后介绍渐近有效性的一些定义.

二、有关随机序列的收敛定理

定理 1.1 若 $X_1, \cdots, X_n, \cdots$ 为 *iid*. 随机序列. 则

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to 0 \text{ a. s.} \Leftrightarrow EX_1$$

存在, 且为 0.

这是有名的 A. H. Kolmogorov 定理. 一般的概率论书上都有证明. 这里证明从略.

定理 1.2(Helly-Bray 定理) 若 p 维分布函数序列 $\{P_n(x)\}$ 依分布收敛于分布函数 $P_0(x)$, $g(x)$ 是 R^p 上的连续函数, 且关于 $\{P_n(x)\}$ 一致可积, 即对任意 $\varepsilon>0$, 在 R^p 中存在区间 $[a, b]$, 使

$$\int_{[a,b]^c} |g(x)|dP_n(x) \leqslant \varepsilon, \quad ([a, b]^c = R^p - [a, b])$$

对 n 一致成立, 则有

$$\int_{R^p} g(x)dP_n(x) \to \int_{R^p} g(x)dP_0(x) \quad (n\to\infty). \tag{1.8}$$

证明从略, 可参见[1].

注 1. 可以证明, $g(x)$ 一致可积也是 (1.8) 式成立的必要条件.

定理 1.3 设 $\{\xi_n\}$、$\{\varepsilon_n\}$ 是一串 p 维随机向量, 并且有 ξ_n 按分布收敛于 ξ, $\varepsilon_n \xrightarrow{P} 0(n\to\infty)$ 又 $\{\eta_n\}$ 是一串 $p\times p$ 维随机矩阵, 且 $\eta_n \xrightarrow{P} a(n\to\infty)$, 其中 a 为常数矩阵, 则当 $n\to\infty$ 时,

$\eta_n\xi_n+\varepsilon_n$ 按分布收敛于 $a\xi$.

证明从略，可参见[1]，留给读者作为习题.

三、母函数与 Chernoff 定理

设 X 是任一随机变量，则称 $M(t)\triangleq Ee^{tX}$ 为 X 的母函数.

因为 $e^{tx}>0$，所以 $M(t)$ 总是存在的，但可能是 $+\infty$，又因为

$$Ee^{tX}=\int_{x\leqslant 0}e^{tx}dP+\int_{x>0}e^{tx}dP,$$

利用单调收敛定理知，$M(\pm\infty)=\lim\limits_{t\to\pm\infty}M(t)$ 存在，故我们可视 $M(t)$ 为在 $\bar{R}=[-\infty,\ \infty]$ 上定义的函数.

记 $T=\{t:M(t)<\infty,\ t\in\bar{R}\}$;

$$a=\inf\{t:t\in T\}\quad (\text{可以是}-\infty);$$

$$b=\sup\{t:t\in T\}\quad (\text{可以是}+\infty).$$

母函数有如下性质:

1) 若 $a<b$，则对任意 $t\in(a,\ b)$，有

$E|X^k|e^{tX}<\infty$，且 $M^{(k)}(t)=E\{X^ke^{tX}\}\quad (k=1,\ 2,\ \cdots)$.

证 对任意 $t_0\in(a,\ b)$，存在 $\delta>0$，使得 $t_0\pm\delta\in(a,\ b)$. 于是 $E(e^{(t_0-\delta)X}+e^{(t_0+\delta)X})<\infty$. 对任意正整数 k，总存在充分大的 M，使

$$|x^k|e^{t_0x}\leqslant e^{(t_0-\delta)x}+e^{(t_0+\delta)x}\quad (\text{当 } |x|\geqslant M).$$

从而

$$E|X^k|e^{t_0X}\leqslant\int_{|x|<M}|x|^ke^{t_0x}dP+E(e^{(t_0-\delta)X}+e^{(t_0+\delta)X})<\infty.$$

于是再利用第一章定理 2.2，性质 1 即得证.

2) 当 X 不退化为零时，$M(t)$ 为 T 上的严凸函数，有唯一极小值点 t^*，而且

当 $EX\leqslant 0$ 时，有 $\quad t^*\geqslant 0$;

当 $EX\geqslant 0$ 时，有 $\quad t^*\leqslant 0$;

当 $P[X>0]=0$ 时，$t^*=+\infty$;

当 $P[X<0]=0$ 时，$t^*=-\infty$;

当 $P(X>0)>0$ 且 $P(X<0)>0$ 时, t^* 有限.

证 当 $P(X=0)<1$ 时, $M^{(2)}(t)=E(X^2e^{tX})>0$, 这证明了 $M(t)$ 的严凸性, 并由此可知, $M(t)$ 有唯一极小值点 t^*. 如果 $T=\{0\}$, 显然 $t^*=0$. 又若 $0\in(a, b)$, 此时 $M'(0)=EX$, 此时当 $EX=0$ 时, $t^*=0$. 若 $EX<0$, $M(t)$在 $t=0$ 处下降, 故$t^*>0$. 同理,若 $EX>0$, $M(t)$在 $t=0$ 处上升, $t^*<0$. 若 $a=0$, $b>0$. 若 $EX<0$, $M'(0+)=EX<0$, 故 $t^*>0$; 当 $EX\geqslant 0$, 因为当 $t\geqslant 0$ 时,

$$\int xe^{tx}dP\geqslant\int x\,dP=EX\geqslant 0,$$

即有 $M'(t)\geqslant 0$, 所以 $t^*=0$, 同理可分析 $b=0$, $a<0$ 时的情况.

若 $P[X>0]=0$, X 不退化为 0, 那么 $M(t)=\int_{x<0}e^{xt}dP$ 为 t 的严格单调下降函数, 故 $t^*=+\infty$. 同理, 若 $P[X<0]=0$, 有 $t^*=-\infty$.

又若 $P(X>0)>0$, $P(X<0)>0$, 当 $t\to\infty$ 时,有

$$M(t)\geqslant\int_{x>0}e^{tx}dP\to\infty,$$

而当 $t\to-\infty$ 时, 有 $M(t)\geqslant\int_{x<0}e^{tx}dP\to\infty$, 但因 $M(0)=1$, 故 t^* 必为有限值.

3) 假设存在 $\delta>0$, 使得 $M(t)<\infty$, 当 $|t|<\delta$, 且 $EX=0$, $\sigma^2=EX^2>0$. 则当 $\varepsilon\to 0$ 时, 有

$$1>\inf_t e^{\pm\varepsilon t}M(t)=e^{-\frac{\varepsilon^2}{2\sigma^2}+O(\varepsilon^3)}. \tag{1.9}$$

证 因 $EX=0$, $\sigma^2=EX^2>0$, 故存在 $\varepsilon>0$, 使 $P(X>\varepsilon)>0$, $P(X<-\varepsilon)>0$. 记 $Z_\varepsilon=X+\varepsilon$, $M_\varepsilon(t)=Ee^{Z_\varepsilon t}=e^{t\varepsilon}Ee^{Xt}$, 则因 $P[Z_\varepsilon<0]>0$, $P[Z_\varepsilon>0]>0$. $EZ_\varepsilon=\varepsilon>0$, 于是由 2)知, 存在唯一有限数 $t_\varepsilon<0$, 使

$$\inf_t M_\varepsilon(t)=M_\varepsilon(t_\varepsilon)=e^{\varepsilon t_\varepsilon}Ee^{Xt_\varepsilon}<M_\varepsilon(0)=1.$$

下面证：当 $\varepsilon \downarrow 0$ 时，$t_\varepsilon \to 0$.

事实上，若 $t_\varepsilon \to 0(\varepsilon \to 0)$ 不成立，则存在 $\varepsilon_n \downarrow 0$，但 $t_{\varepsilon_n} \to t_0 < 0$. 注意 t_{ε_n} 为 $M_{\varepsilon_n}(t)$ 的极小值点，故有 $M'_{\varepsilon_n}(t_{\varepsilon_n}) = 0$. 于是有

$$0 = \lim_{n\to\infty} M'_{\varepsilon_n}(t_{\varepsilon_n}) = \lim_{n\to\infty} E[(X+\varepsilon_n)e^{(X+\varepsilon_n)t_{\varepsilon_n}}]$$

$$\leqslant E[\limsup_{n\to\infty}(X+\varepsilon_n)e^{(X+\varepsilon_n)t_{\varepsilon_n}}] = E(Xe^{Xt_0})$$

$$= M'(t_0) < 0.$$

得出矛盾. 注意：上式由已知条件容易验证利用控制收敛定理是合理的，最后一步是因为 $M'(0) = EX = 0$，所以 $t=0$ 为 $M(t)$ 唯一极小值点，而 $t_0<0$，故 $M'(t_0)<0$.

再注意 $M''(0) = EX^2 = \sigma^2$, $M'(0) = EX = 0$，于是利用 Taylor 展开式，有

$$0 = E[Z_\varepsilon e^{Z_\varepsilon t_\varepsilon}] = \varepsilon e^{\varepsilon t_\varepsilon} M(t_\varepsilon) + e^{\varepsilon t_\varepsilon} M'(t_\varepsilon)$$
$$= \varepsilon(1+\varepsilon t_\varepsilon + O(\varepsilon^2 t_\varepsilon^2))[1+O(t_\varepsilon^2)] + (1+\varepsilon t_\varepsilon + O(\varepsilon^2 t_\varepsilon))$$
$$(M''(0)t_\varepsilon + O(t_\varepsilon^2)) = \varepsilon + \sigma^2 t_\varepsilon + O(\varepsilon^2 t_\varepsilon) + O(t_\varepsilon^2).$$

从而可知，当 $\varepsilon \downarrow 0$ 时，t_ε 与 ε 为同阶无穷小，且

$$t_\varepsilon = -\frac{\varepsilon}{\sigma^2} + O(\varepsilon^2).$$

于是

$$M(t_\varepsilon) = 1 + M'(0)t_\varepsilon + \frac{1}{2}M''(0)t_\varepsilon^2 + O(t_\varepsilon^3)$$

$$= 1 + \frac{\varepsilon^2}{2\sigma^2} + O(\varepsilon^3) = e^{\frac{\varepsilon^2}{2\sigma^2} + O(\varepsilon^3)}. \tag{1.10}$$

由此便推出 $M(t_\varepsilon)e^{\varepsilon t_\varepsilon} = e^{\frac{\varepsilon^2}{2\sigma^2} - \frac{\varepsilon^2}{\sigma^2} + O(\varepsilon^3)} = e^{-\frac{\varepsilon^2}{2\sigma^2} + O(\varepsilon^3)}$.

同样可证 $\inf_t e^{-\varepsilon t}M(t) = e^{-\frac{\varepsilon^2}{2\sigma^2} + O(\varepsilon^3)}$.

4) 设 $X_1, \cdots, X_n$ 为 n 个互相独立的随机变量，记 $M_i(t) = Ee^{tX_i}$ $(i=1, \cdots, n)$，则 $S_n = \sum_1^n x_i$ 的母函数 $M(t) = \prod_{i=1}^n M_i(t)$. 这利用 $X_1, \cdots, X_n$ 的相互独立性即可得到.

定理 1.4(Chernoff) 设 $X_1, X_2, \cdots$ 为 *iid.* 随机序列. 记

$M(t)=Ee^{X_1t}$, $S_n=\sum_{i=1}^{n}X_i$, $m=\inf_t M(t)$，则

i) 如果 $E(X_1)\geqslant 0$，则

$$P(S_n\leqslant 0)\leqslant m^n; \tag{1.11}$$

如果 $E(X_1)\leqslant 0$，则

$$P(S_n\geqslant 0)\leqslant m^n. \tag{1.12}$$

ii) 若存在 $\delta>0$，使当 $|t|<\delta$ 时，$M(t)<\infty$. 那么当 $E(X_1)\neq 0$ 时，有 $m<1$.

证 i) 因为对任意的 $t\leqslant 0$，皆有

$$P(S_n\leqslant 0)\leqslant\int_{S_n<0}e^{tS_n}dP\leqslant Ee^{tS_n}=(M(t))^n.$$

对上式两边在 $t\leqslant 0$ 范围内取极小值，即得

$$P(S_n\leqslant 0)\leqslant(\inf_{t<0}M(t))^n.$$

又由母函数性质 2)，当 $EX_1\geqslant 0$ 时，$M(t)$ 在 $t\leqslant 0$ 时达到极小值，故 $\inf_{t<0}M(t)=\inf_t M(t)=m$. 因此 (1.11) 式得证. 用同样方法可证(1.12)式.

ii) 由母函数的性质 1) 和 2) 可知，$M'(0)=EX_1\neq 0$，故 $M(t)$ 的唯一极小值点 $t^*\neq 0$，从而 $\inf_t M(t)=M(t^*)<M(0)=1$. 定理证毕.

注 原定理结论还有：若 $m>0$，则对任意的 $0<\varepsilon<m$，有 $\lim_{n\to\infty}(m-\varepsilon)^n/P(S\leqslant 0)=\lim_{n\to\infty}(m-\varepsilon)^n/P(S_n\geqslant 0)=0$. 因为这结果本书未用到，加之证明冗长，故略去.

四、Kullbuck-Leibler 信息函数

定义 1.6 设随机向量 X 的样本空间为$(\mathscr{X},\mathscr{B}_x,P_\theta,\theta\in\Theta)$，其中参数空间 Θ 为 R^p 中的开域，P_θ 关于 $\mathscr{B}_x$ 上的 σ-有限测度 μ 的概率密度 $f(x,\theta)$ 对所有的 $\theta\in\Theta$ 有共同支撑，即集合$\{x:f(x,\theta)>0\}$与 θ 无关. 对任意的 φ、$\theta\in\Theta$，令

$$I(\varphi,\theta)=E_\varphi\log(f(x,\varphi)/f(x,\theta)), \tag{1.13}$$

称 $I(\varphi, \theta)$ 为 X 的 Kullbuck-Leibler 信息函数，简称为 K-L 信息函数. 有时为明确起见，也记为 $I^X(\varphi, \theta)$.

K-L 信息函数有以下性质：

1° 对任意的 θ、$\varphi\in\Theta$, $I(\varphi, \theta)$ 皆有意义，且与 μ 选择无关.

2° $I(\varphi, \theta)\geqslant 0$, 而且

$$I(\varphi, \theta)=0 \Leftrightarrow f(x, \varphi)=f(x, \theta) \quad \text{a. s. } \mu. \tag{1.14}$$

3° 若随机向量 X 与 Y 相互独立，它们关于 μ_1, μ_2 的概率密度 $f_X(x, \theta)$, $f_Y(y, \theta)$, $\theta\in\Theta$, 有共同的支撑，则有

$$I^{(X,Y)}(\varphi, \theta)=I^X(\varphi, \theta)+I^Y(\varphi, \theta), \quad \varphi、\theta\in\Theta, \tag{1.15}$$

又当 $X_1, \cdots, X_n$ 为 iid. 随机变量，其分布族有共同支撑，则

$$I^{(X_1,\cdots,X_n)}(\varphi, \theta)=nI^{X_1}(\varphi, \theta). \tag{1.16}$$

证 1° 记 $(h(x))^+$、$(h(x))^-$ 分别为函数 $h(x)$ 的正部和负部. 于是

$$\left[\log\frac{f(x, \varphi)}{f(x, \theta)}\right]^-=\left[\log\frac{f(x, \theta)}{f(x, \varphi)}\right]^+\leqslant\frac{f(x, \theta)}{f(x, \varphi)},$$

而 $E_\varphi\left\{\frac{f(X, \theta)}{f(X, \varphi)}\right\}=\int f(x,\theta)d\mu=1$, 故 $E_\varphi\left[\log\frac{f(X, \varphi)}{f(X, \theta)}\right]^-<\infty$,

$$\begin{aligned}I(\varphi, \theta)=E_\varphi\log\frac{f(x, \varphi)}{f(x, \theta)}&=E_\varphi\left[\log\frac{f(X, \varphi)}{f(X, \theta)}\right]^+\\&\quad-E_\varphi\left[\log\frac{f(X, \varphi)}{f(X, \theta)}\right]^-\end{aligned}$$

有意义. 它与 μ 的选择无关，可由定义直接推出.

2° 因为 $-\log y$ 是 y 的严凸函数，利用 Jensen 不等式（第一章定理 3.7）. 有

$$\begin{aligned}I(\varphi, \theta)=E_\varphi\left[-\log\frac{f(X, \theta)}{f(X, \varphi)}\right]&\geqslant-\log E_\varphi\left\{\frac{f(X, \theta)}{f(X, \varphi)}\right\}\\&=-\log 1=0,\end{aligned}$$

而且上式等号成立 $\Leftrightarrow f(X,\theta)/f(X,\varphi)=1$, a.s. $P_\varphi\Leftrightarrow f(x, \theta)=f(x, \varphi)$ a. s. P_φ. 但注意到 P_θ, $\theta\in\Theta$ 有共同支撑，而 μ 为在此支撑空间上定义的 σ-有限测度，故有 $P_\theta\equiv P_\varphi\equiv\mu$, 对任意的 θ、$\varphi\in\Theta$. 于是 (1.14) 式得证. 性质 2° 说明了 $I(\varphi, \theta)$ 用来作为区分真假参数值 φ、θ 的标志是合理的.

3° 由 X 与 Y 的独立性质，(X, Y)的概率密度为

$$f^{(X,Y)}(x, y, \theta)=f_X(x, \theta)f_Y(y, \theta).$$

于是

$$\begin{aligned}I^{(X,Y)}(\varphi, \theta)&=E_\varphi \log\left[\frac{f_X(X, \varphi)f_Y(Y, \varphi)}{f_X(X, \theta)f_Y(Y, \varphi)}\right]\\&=E_\varphi \log\left[\frac{f_X(X, \varphi)}{f_X(X, \theta)}\right]\\&\quad+E_\varphi \log\left[\frac{f_Y(Y, \varphi)}{f_Y(Y, \theta)}\right]\\&=I^{(X)}(\varphi, \theta)+I^{(Y)}(\varphi, \theta).\end{aligned}$$

于是(1.15)式得证. (1.16)是(1.15)的自然结果.

下面我们来估计一下，当 $\theta\to\varphi$ 时，$I(\varphi, \theta)$ 的近似值. 根据 K-L 信息函数性质 2°，我们不妨设 $\mu(x:f(x, \theta)\neq f(x, \varphi))>0$ 当 $\varphi\neq\theta$. 先给出下面的引理.

引理 1.1 设对任意的 $\varphi\in\Theta$，存在 φ 的邻域 $V_\varphi\subset\Theta$，使得当 $\theta\in V_\varphi$ 时，

$$\left|\frac{f(x, \theta)}{f(x, \varphi)}-1\right|\leqslant A(x, \varphi)|\varphi-\theta|. \tag{1.17}$$

$\frac{\partial \log f(x, \varphi)}{\partial\varphi}\triangleq Z(x, \varphi)$存在，有限，对每个 x 关于 φ 连续，且 $E_\varphi[Z(X, \varphi)Z'(X, \varphi)]\triangleq I(\varphi)>0$.

$$|Z(x, \theta)|\leqslant B(x, \varphi) \quad (\text{当 } \theta\in V_\varphi \text{ 时}), \tag{1.18}$$

其中

$$E_\varphi A(X, \varphi)<\infty,\ E_\varphi B^2(X, \varphi)<\infty,$$
$$E_\varphi A(X, \varphi)B^2(X, \varphi)<\infty.$$

这里 $\frac{\partial}{\partial\varphi}=\left(\frac{\partial}{\partial\varphi_1}, \cdots, \frac{\partial}{\partial\varphi_p}\right)'$ 表示微分算子. 则有

i) $E_\varphi Z(X, \varphi)=0$; (1.19)

ii) $E_\varphi Z(X, \theta)=-I(\varphi)(\theta-\varphi)(1+o(1))$; (1.20)

iii) $I(\varphi)$是 φ 的连续函数;

iv) $I(\theta, \varphi)=\frac{1}{2}(\varphi-\theta)'I(\varphi)(\varphi-\theta)(1+o(1))$; (1.21)

v) $E_\varphi\left[\log\frac{f(X,\varphi)}{f(X,\theta)}\right]^2=(\varphi-\theta)'I(\varphi)(\varphi-\theta)(1+o(1))$.

$$(1.21)'$$

此处 $o(1)$ 是指 $\theta\to\varphi$ 而言.

证 i) 记 $Z(x,\varphi)=(Z_1(x,\varphi),\cdots,Z_P(x,\varphi))'$,

$\delta_i=(\underbrace{0,\cdots,0}_{i-1\text{个}},1,0,\cdots,0)'$. 利用假设 (1.17) 式可知, 用控制收敛定理, 有

$$\begin{aligned}E_\varphi Z_i(X,\varphi)&=E_\varphi\left[\frac{\partial f(x,\varphi)}{\partial\varphi_i}\Big/f(x,\varphi)\right]\\&=\int\lim_{\Delta\varphi_i\to0}\frac{f(x,\varphi+\delta_i\Delta\varphi_i)-f(x,\varphi)}{\Delta\varphi_i}d\mu\\&=\lim_{\Delta\varphi_i\to0}\frac{1}{\Delta\varphi_i}\int(f(x,\varphi+\delta_i\Delta\varphi_i)-f(x,\varphi))d\mu\\&=\lim_{\Delta\varphi_i\to0}\frac{1-1}{\Delta\varphi_i}=0\end{aligned}$$

$(i=1,\cdots,P)$. 从而有 $E_\varphi Z(X,\varphi)=0$.

iii) 由假设(1.17)与(1.18)式, 当 $\theta\in V_\varphi$ 时, 有

$$f(x,\theta)\leqslant f(x,\varphi)(1+A(x,\varphi)|\varphi-\theta|);\qquad(1.22)$$

$$|Z(x,\theta)Z'(x,\theta)|f(x,\theta)\leqslant B^2(x,\varphi)(1+A(x,\varphi)|\varphi-\theta|)f(x,\varphi).\qquad(1.23)$$

而且 $E_\varphi B^2(X,\varphi)<\infty$, $E_\varphi B^2(X,\varphi)A(X,\varphi)<\infty$, 再利用控制收敛定理, 极限号与积分号可交换, 从而有 $\lim\limits_{\theta\to\varphi}I(\theta)=I(\varphi)$. 于是 iii)得证.

ii) $$\begin{aligned}E_\varphi[Z(X,\theta)]&=\int Z(x,\theta)f(x,\varphi)d\mu\\&=\int Z(x,\theta)(f(x,\varphi)-f(x,\theta))d\mu\\&\triangleq\int Z(x,\theta)d\mu\int_0^1[Z(x,\varphi+t(\theta-\varphi))]'\\&\quad\times(\varphi-\theta)f(x,\varphi+t(\theta-\varphi))dt.\end{aligned}\qquad(1.24)$$

再利用(1.22)及(1.18), 就有

$$|\dot{Z}(x,\ \theta)(\dot{Z}(x,\ \varphi+t(\varphi-\theta)))'|f(x,\ \varphi+t(\varphi-\theta))$$
$$\leqslant B^2(x,\ \varphi)(1+A(x,\ \varphi)|\varphi-\theta|)f(x,\ \varphi). \qquad (1.25)$$

如同证明 $I(\theta)$ 连续时一样，用控制收敛定理可推得(1.24)最后一个等式的右端为 $I(\varphi)(\varphi-\theta)(1+o(1))$，故 ii)得证.

至于 iv)的证明与 ii)基本相同:

$$I(\theta,\ \varphi)=\int\left(\log\frac{f(x,\ \theta)}{f(x,\ \varphi)}\right)f(x,\ \theta)d\mu$$
$$=\int f(x,\ \theta)d\mu\int_0^1(Z(x,\ \varphi+t(\theta-\varphi)))'(\theta-\varphi)dt$$
$$=\int_0^1 dt\int(Z(x,\ \varphi+t(\theta-\varphi)))'(\theta-\varphi)f(x,\ \theta)d\mu$$
$$=-\int_0^1(I(\varphi)(1-t)(\varphi-\theta)(1+o(1)))'dt(\theta-\varphi)$$
$$=\frac{1}{2}(\theta-\varphi)'I(\varphi)(\theta-\varphi)(1+o(1)).$$

其中用到

$$\int Z(x,\ \varphi+t(\theta-\varphi))f(x,\ \theta)d\mu$$
$$=-I(\varphi)(1-t)(\varphi-\theta)(1+o(1)),$$

它的证明与 ii)的证明相同. 至于 v)的证明与 iv)相同，而且更简单，故从略.

注 此引理 $Z(x,\ \varphi)$ 的存在与连续的条件可减弱为: 对每个 $\varphi\in\Theta$, 对几乎所有的 x 在 φ 的某个邻域内存在且连续，引理结论仍然成立.

为了区分不同参数，我们还可以用密度函数的距离作标志. 也即

$$d(\theta,\ \varphi)=\int|f(x,\ \theta)-f(x,\ \varphi)|d\mu. \qquad (1.26)$$

它的好处是: 并不要求 $f(x,\ \theta)$, $\theta\in\Theta$ 有共同的支撑，因此适应范围更广. 但是它没有 3° 那样好的性质，从而不便于使用. 下面给出一个引理，进一步说明 $d(\theta,\ \varphi)$ 的含义.

引理 1.2

$$d(\theta,\ \varphi)=2\sup_{A\in\mathscr{B}_x}|P_\theta(A)-P_\varphi(A)|, \qquad (1.27)$$

此处 $\mathscr{B}_x$ 为 X 值域空间 $\mathscr{X}$ 上的 Borel σ-域.

证

$$
\begin{aligned}
&2|P_\theta(A)-P_\varphi(A)| \\
&=|P_\theta(A)-P_\varphi(A)|+|P_\theta(A^c)-P_\varphi(A^c)| \\
&\leqslant\int_A|f(x,\theta)-f(x,\varphi)|d\mu \\
&\quad+\int_{A^c}|f(x,\theta)-f(x,\varphi)|d\mu=d(\theta,\varphi).
\end{aligned}
$$

另一方面，令 $A_0=\{x:f(x,\theta)>f(x,\varphi)\}$.

$$
\begin{aligned}
d(\theta,\varphi)&=\int_{A_0}(f(x,\theta)-f(x,\varphi))d\mu \\
&\quad+\int_{A_0^c}(f(x,\varphi)-f(x,\theta))d\mu \\
&=2|P_\theta(A_0)-P_\varphi(A_0)| \\
&\leqslant 2\sup_{A\in\mathscr{B}_x}|P_\theta(A)-P_\varphi(A)|.
\end{aligned}
$$

故引理得证.

五、Neyman-Pearson 基本引理

设 X 是一个多维随机变量，考虑如下假设检验：零假设 H_0: $X\sim P_0$, 对立假设 H_1: $X\sim P_1$. P_0, P_1 为关于 σ-有限测度 μ(可取 P_0+P_1)的密度，记为 $f_0(x)$, $f_1(x)$, α 是任意给定的显著性水平($0\leqslant\alpha\leqslant1$)，则

(i) 必有一检验函数 $\varphi(x)$ (已知 x 拒绝 H_0 的概率) 和常数 k, 满足:

(1) $$\varphi(x)=\begin{cases}1, & \text{当 } f_1(x)>kf_0(x); \\ 0, & \text{当 } f_1(x)<kf_0(x).\end{cases} \tag{1.28}$$

(2) $$E_0\{\varphi(X)\}=\alpha. \tag{1.29}$$

(ii) 满足(i)的检验 $\varphi(x)$ 一定是显著水平为 α 的最大功效检验，即对任意的检验 $\tilde{\varphi}(x)$, 若满足 $E_0\tilde{\varphi}(X)\leqslant\alpha$, 必有 $E_1\tilde{\varphi}\leqslant E_1\varphi$. $E_1\varphi$ 称为 φ 的功效.

(iii) 如果 φ 是显著性水平 α 上的最大功效检验，则 α 必定对

某一常数 k 满足(i)中的 (1)、(2) 式，除非存在一个检验 φ_0，满足 $E_1\varphi_0=1$，$E_0\varphi_0<\alpha$.

证明从略，参见[3].

§2 相合估计

一、一般必要条件

上节我们已经给出了相合估计的定义. 本节我们将给出参数 θ 存在相合估计的必要条件，同时将讨论线性模型中的最小二乘估计和一般概率密度函数估计的相合性问题. §1 中所给出的概率空间序列 $(\mathscr{X}_n,\ \mathscr{B}_{x_n},\ P_{n,\theta},\ \theta\in\Theta)$，$dP_{n,\theta}(x_n)=p_n(x_n,\ \theta)d\mu_n$，这些记号及相应的假设条件在下面仍然要运用.

定理 2.1 θ 存在相合估计的必要条件是：对任意 $\theta_1\neq\theta_2(\in\Theta)$，有 $\lim\limits_{n\to\infty}d_n(\theta_1,\ \theta_2)=2$. 此处

$$d_n(\theta_1,\ \theta_2)=\int|p_n(x_n,\ \theta_1)-p_n(x_n,\ \theta_2)|d\mu_n.$$

证 如果 $\hat{\theta}_n$ 是 θ 的相合估计，记 $\varepsilon=|\theta_1-\theta_2|/2$. 令

$$\begin{aligned}A_n&=\{x_n:|\hat{\theta}_n(x_n)-\theta_1|<\varepsilon\};\\B_n&=\{x_n:|\hat{\theta}_n(x_n)-\theta_2|<\varepsilon\}.\end{aligned}\tag{2.1}$$

显然 A_n 与 B_n 互不相交. 因为 $\hat{\theta}_n$ 是 θ 的相合估计，故

$$\lim_{n\to\infty}P_{n,\theta_1}(A_n)=1,\qquad\lim_{n\to\infty}P_{n,\theta_2}(A_n)=0.\tag{2.2}$$

从而根据上节引理 1.2，

$$\begin{aligned}d_n(\theta_1,\ \theta_2)&=2\sup_{A\in\mathscr{B}_{x_n}}|P_{n,\theta_1}(A)-P_{n,\theta_2}(A)|\\&\geqslant2(P_{n,\theta_1}(A_n)-P_{n,\theta_2}(A_n))\to2\quad(n\to\infty).\end{aligned}$$

但由 $d_n(\theta_1,\ \theta_2)$ 的定义知，总有 $2\geqslant d_n(\theta_1,\ \theta_2)$. 故得

$$\lim_{n\to\infty}d_n(\theta_1,\ \theta_2)=2.\tag{2.3}$$

定理 2.2 如果 $p_n(x_n,\ \theta)$，$\theta\in\Theta$ 有共同支撑. 则 θ 具有相合估计的必要条件是 K-L 信息函数对所有的 $\theta_1\neq\theta_2(\in\Theta)$ 有 $I_n(\theta_1,\ \theta_2)\to\infty\ (n\to\infty)$.

证　令 $Y_n=p_n(X_n,\ \theta_2)/p_n(X_n,\ \theta_1)$，显然 $E_{\theta_1}(Y_n)=1$. 记

$$Y_n^+=\max(Y_n-1,\ 0);\quad Y_n^-=\max(1-Y_n,\ 0).$$

因为 $Y_n-1=Y_n^+-Y_n^-$, $E_{\theta_1}(Y_n)=1$, 故有 $E_{\theta_1}Y_n^+=E_{\theta_1}Y_n^-$,

$$d_n(\theta_1,\ \theta_2)=\int|p_n(x_n,\ \theta_2)-p_n(x_n,\ \theta_1)|d\mu_n$$

$$=\int|Y_n-1|p_n(x_n,\ \theta_1)d\mu_n=E_{\theta_1}(Y_n^++Y_n^-),$$

$$\begin{aligned}I_n(\theta_1,\ \theta_2)&=-E_{\theta_1}\log Y_n\\&=-\int_{y_n>1}\log y_n\,dP_{n,\theta_1}-\int_{0<y_n<1}\log y_n\,dP_{n,\theta_1}\\&=-\{E_{\theta_1}\log(1+Y_n^+)+E_{\theta_1}\log(1-Y_n^-)\}\\&\geqslant-\log E_{\theta_1}(1+Y_n^+)-\log E_{\theta_1}(1-Y_n^-).\end{aligned}\tag{2.4}$$

其中，最后一步利用了 Jenson 不等式. 又由于

$$E_{\theta_1}Y_n^+=E_{\theta_1}Y_n^-=\frac{1}{2}d_n(\theta_1,\ \theta_2).$$

将上式代入(2.4)式，并注意 $\lim\limits_{n\to\infty}d_n(\theta_1,\ \theta_2)=2$ (由定理 2.1)，令 $n\to\infty$，有

$$\begin{aligned}\liminf_{n\to\infty}I_n(\theta_1,\ \theta_2)&\geqslant-\lim_{n\to\infty}\log\left(1+\frac{1}{2}d_n(\theta_1,\ \theta_2)\right)\\&-\lim_{n\to\infty}\log\left(1-\frac{1}{2}d_n(\theta_1,\ \theta_2)\right)=-\log 2+\infty=\infty.\end{aligned}\tag{2.5}$$

至此定理证毕.

二、线性模型中 LS 估计的相合性

下面研究线性模型中回归系数估计的相合性问题. 考虑下面线性回归模型:

$$y_i=x_i'\beta+\varepsilon_i\quad(i=1,\ 2,\ \cdots).\tag{2.6}$$

此处，β 是回归系数，为 p 维未知向量，$\{x_i,\ i=1,\ 2,\ \cdots\}$ 为已知的 p 维向量序列，假定存在 n_0，使 $\sum\limits_{i=1}^{n_0}x_ix_i'>0$(正定)，$\varepsilon_1,\ \varepsilon_2,\cdots$ 为 *iid*. 随机误差序列，满足条件

$$E\varepsilon_1=0,\ E\varepsilon_1^2=\sigma^2<\infty,$$

而 $y_1, y_2, \cdots$ 为可观察的随机变量序列. 在这些假定之下, 我们有如下关于回归系数最小二乘估计的相合性的定理:

定理 2.3 i) β 的最小二乘估计 $\hat{\beta}_n = S_n^{-1}X_n'Z_n (n \geqslant n_0)$ 为相合估计的充要条件是 $S_n^{-1} \to 0 (n \to \infty)$. 这里

$$X_n' = (x_1, \cdots, x_n), \quad Z_n = (y_1, \cdots, y_n)';$$

$$S_n = X_n'X_n = \sum_1^n x_i x_i'.$$

ii) $\hat{\sigma}_n^2 = \dfrac{1}{n-P} Z_n'(I_n - X_nS_n^{-1}X_n')Z_n \quad (n \geqslant n_0)$ 是 σ^2 的相合估计.

证 i) 充分性: 由第二章 § 6(6.13)式知

$$\mathrm{Var}\,\hat{\beta}_n = S_n^{-1}\sigma^2. \tag{2.7}$$

由假设条件 $S_n^{-1} \to 0 (n \to \infty)$, 从而

$$E(\|\hat{\beta}_n - \beta\|^2) = \mathrm{tr}S_n^{-1}\sigma^2 \to 0 \quad (n \to \infty).$$

故 $\hat{\beta}_n$ 均方收敛于 β, 从而 $\hat{\beta}_n \xrightarrow{P} \beta \ (n \to \infty)$. 充分性得证.

i) 必要性: 我们先考虑 β 的第一个分量的 LS 估计. 令

$$T_i = (x_{i2}, \cdots, x_{iP})' \quad (i = 1, 2, \cdots),$$

于是有

$$X_n' = \begin{bmatrix} x_{11}, & \cdots, & x_{n1} \\ T_1, & \cdots, & T_n \end{bmatrix}, \quad S_n = X_n'X_n = \begin{bmatrix} \sum\limits_{i=1}^n x_{i1}^2 & K_n \\ K_n' & H_n \end{bmatrix}.$$

其中 $K_n = \sum\limits_{i=1}^n x_{i1}T_i'$, $H_n = \sum\limits_{i=1}^n T_iT_i'$. 由求逆公式,

$$S_n^{-1} =$$

$$\begin{bmatrix} \left(\sum\limits_1^n x_{i1}^2 - K_nH_n^{-1}K_n'\right)^{-1} & -\left(\sum\limits_1^n x_{i1}^2 - K_nH_n^{-1}K_n^1\right)^{-1} K_nH_n^{-1} \\ -\left(\sum\limits_1^n x_{i1}^2 - K_nH_n^{-1}K_n'\right)^{-1} H_n^{-1}K_n' & * \end{bmatrix} \tag{2.8}$$

因 $\hat{\beta}_n = S_n^{-1}X_n'Z_n = S_n^{-1}X_n'(X_n\beta + \varepsilon(n)) = \beta + S_n^{-1}X_n'\varepsilon(n)$, 此处 $\varepsilon(n) = (\varepsilon_1, \cdots, \varepsilon_n)'$. 由假设条件 $\hat{\beta}_n \xrightarrow{P} \beta$, 则有

$$S_n^{-1}X_n'\varepsilon(n) \xrightarrow{P} 0 \quad (n \to \infty). \tag{2.9}$$

注意

$$'X'_n\varepsilon(n)=\begin{bmatrix}\sum_{i=1}^{n}x_{i1}\varepsilon_i\\ \sum_{i=1}^{n}T_i\varepsilon_i\end{bmatrix}, \tag{2.10}$$

由(2.8)、(2.10)推得 $S_n^{-1}X'_n\varepsilon(n)$ 的第一个分量是

$$\left(\sum_1^n x_{i1}^2-K_nH_n^{-1}K'_n\right)^{-1}\sum_{i=1}^{n}(x_{i1}-K_nH_n^{-1}T_i)\varepsilon_i. \tag{2.11}$$

记 $W_n=\sum_{i=1}^{n}x_{i1}^2-K_nH_n^{-1}K'_n$, 不难推得

$$\begin{aligned}\sum_{i=1}^{n}(x_{i1}-K_nH_n^{-1}T_i)^2&=\sum_{i=1}^{n}x_{i1}^2-2\sum_{i=1}^{n}x_{i1}K_nH_n^{-1}T_i\\&\quad+K_nH_n^{-1}\sum_{i=1}^{n}T_iT'_iH_n^{-1}K'_n\\&=\sum_{i=1}^{n}x_{i1}^2-K_nH_n^{-1}K'_n=W_n.\end{aligned} \tag{2.12}$$

注 意: $H_n=\sum_{i=1}^{n}T_iT'_i$, $K_n=\sum_{i=1}^{n}x_{i1}T'_i$, 于是由最小二乘法易知, 对任意 p 维向量 α, 皆有

$$\sum_{i=1}^{n}(x_{i1}-T'_i\alpha)^2\geqslant\sum_{i=1}^{n}(x_{i1}-T'_iH_n^{-1}K'_n)^2, \tag{2.13}$$

故有

$$\begin{aligned}W_{n+1}&=\sum_{i=1}^{n+1}(x_{i1}-K_{n+1}H_{n+1}^{-1}T_i)^2\\&\geqslant(x_{n+1,1}-T'_{n+1}H_{n+1}^{-1}K'_{n+1})^2\\&\quad+\sum_{i=1}^{n}(x_{i1}-K_nH_n^{-1}T_i)^2\geqslant W_n.\end{aligned} \tag{2.14}$$

从而 $\lim\limits_{n\to\infty}W_n$ 存在. 如果极限值有限, 记为 W_0, 此时 $W_n\leqslant W_0$, $(x_{i1}-K_nH_n^{-1}T_i)^2\leqslant W_0$ $(n=n_0+1,\cdots)$. 故存在自然序列的子列 $\{n_\nu\}$, 使

$$\lim_{\nu\to\infty}(x_{i1}-K_{n_\nu}H_{n_\nu}^{-1}T_i)=d_i\quad(i=1,2,\cdots).$$

显然有
$$\lim_{\nu\to\infty}\sum_{i=1}^{l}(x_{i1}-K_{n_\nu}H_{n_\nu}^{-1}T_i)^2=\sum_{i=1}^{l}d_i^2,$$

$$W_0=\lim_{\nu\to\infty}W_{n_\nu}\geqslant\lim_{\nu\to\infty}\sum_{i=1}^{l}(x_{i1}-K_{n_\nu}H_{n_\nu}^{-1}T_i)^2=\sum_{i=1}^{l}d_i^2.$$

再令 $l\to\infty$, 即有

$$W_0\geqslant\sum_{i=1}^{\infty}d_i^2. \tag{2.15}$$

从(2.13)可知:

$$\sum_{i=1}^{l}(x_{i1}-K_{n_\nu}H_{n_\nu}^{-1}T_i)^2\geqslant\sum_{i=1}^{l}(x_{i1}-K_lH_l^{-1}T_i)^2=W_l. \tag{2.16}$$

$$\lim_{\nu\to\infty}\sum_{i=1}^{l}(x_{i1}-K_{n_\nu}H_{n_\nu}^{-1}T_i)^2=\sum_{i=1}^{l}d_i^2\geqslant W_l. \tag{2.17}$$

令 $l\to\infty$, 即得 $\sum_{i=1}^{\infty}d_i^2\geqslant W_0$. 再由(2.15), 得

$$W_0=\sum_{i=1}^{\infty}d_i^2. \tag{2.18}$$

记 $U_n=\sum_{i=1}^{n}(x_{i1}-K_nH_n^{-1}T_i)\varepsilon_i$, 下面证明

$$\lim_{\nu\to\infty}E\Big(U_{n_\nu}-\sum_{i=1}^{\infty}d_i\varepsilon_i\Big)=0,$$

$$\begin{aligned}
R_{n_\nu}&\triangleq E\Big[U_{n_\nu}-\sum_{i=1}^{\infty}d_i\varepsilon_i\Big]^2\\
&=E\Big[\sum_{i=1}^{n_\nu}(x_{i1}-K_{n_\nu}^1H_{n_\nu}^{-1}T_i-d_i)\varepsilon_i\Big]^2\\
&\quad+E\Big[\sum_{i=n_\nu+1}^{\infty}d_i\varepsilon_i\Big]^2.\\
&=\sigma^2\Big[\sum_{i=1}^{n_\nu}(x_{i1}-K_{n_\nu}H_{n_\nu}^{-1}T_i-d_i)^2+\sum_{i=n_\nu+1}^{\infty}d_i^2\Big]\\
&\leqslant\sigma^2\Big[\sum_{i=1}^{l}(x_{i1}-K_{n_\nu}H_{n_\nu}^{-1}T_i-d_i)^2\\
&\quad+2\sum_{i=l+1}^{n_\nu}(x_{i1}-K_{n_\nu}H_{n_\nu}^{-1}T_i)^2\\
&\quad+2\sum_{i=l+1}^{n_\nu}d_i^2+\sum_{i=n_\nu+1}^{\infty}d_i^2\Big]\\
&\leqslant\sigma^2\Big[\sum_{i=1}^{l}(x_{i1}-K_{n_\nu}H_{n_\nu}^{-1}T_i-d_i)^2+2(W_{n_\nu}-W_l)\\
&\quad+2\sum_{i=l+1}^{n_\nu}d_i^2+\sum_{i=n_\nu+1}^{\infty}d_i^2\Big].
\end{aligned} \tag{2.19}$$

在(2.19)式的两端令$\nu\to\infty$，然后再令$l\to\infty$，得

$$0\leqslant\lim_{\nu\to\infty}\sup R_{n_\nu}\leqslant\lim_{l\to\infty}2\sigma^2\left[W_0-W_l+\sum_{i=l+1}^{\infty}d_i^2\right]=0, \quad (2.20)$$

故$\lim\limits_{\nu\to\infty}R_{n_\nu}=0$. 另一方面，因为

$$E\left(\sum_{i=1}^{\infty}d_i\varepsilon_i\right)^2=\sum_{i=1}^{\infty}d_i^2=W_0>0,$$

所以$P\left[\sum\limits_{i=1}^{\infty}d_i\varepsilon_i=0\right]<1$. 但从$\lim\limits_{\nu\to\infty}R_{n_\nu}=0$，推得

$$W_{n_\nu}^{-1}U_{n_\nu}\xrightarrow{P}\frac{1}{W_0}\sum_{i=1}^{\infty}d_i\varepsilon_i, \quad (\nu\to\infty)$$

又因$\hat{\beta}_n\xrightarrow{P}\beta(n\to\infty)$，由(2.9)和(2.11)式推得

$$W_{n_\nu}^{-1}U_{n_\nu}\xrightarrow{P}0 \quad (\nu\to\infty). \quad (2.21)$$

这与$P\left[\sum\limits_{i=1}^{\infty}d_i\varepsilon_i=0\right]<1$相矛盾. 由此可知，$W_0$不为有限数，即有$\lim\limits_{n\to\infty}W_n=\infty$或$\lim\limits_{n\to\infty}W_n^{-1}=0$. 这便证明了$S_n^{-1}$的第一行第一列元素$\to0$. 同样可证$S_n^{-1}$的其它主对角线上的元素$\to0$. 又因$S_n^{-1}$正定$(n\geqslant n_0)$，从而证得$\lim\limits_{n\to\infty}S_n^{-1}=0$. i)的必要性得证.

再证ii)：因$n\geqslant n_0$时，

$$\begin{aligned}\hat{\sigma}_n^2&=\frac{1}{n-P}(X_n\beta+\varepsilon(n))'(I_n-X_nS_n^{-1}X_n')(X_n\beta+\varepsilon(n))\\&=\frac{1}{n-P}(\varepsilon(n))'(I_n-X_nS_n^{-1}X_n')\varepsilon(n).\end{aligned}$$

而

$$\begin{aligned}&\frac{1}{n-P}E(\varepsilon(n))'X_nS_n^{-1}X_n'\varepsilon(n)\\&=\frac{1}{n-P}E\mathrm{tr}X_nS_n^{-1}X_n'(\varepsilon(n))(\varepsilon(n))'\\&=\frac{\sigma^2}{n-P}\mathrm{tr}X_nS_n^{-1}X_n'=\frac{P}{n-P}\sigma^2\to0 \quad (n\to\infty).\end{aligned} \quad (2.22)$$

从而由上式知

$$\frac{1}{n-P}(\varepsilon(n))'X_nS_n^{-1}X_n'\varepsilon(n)\xrightarrow{P}0 \quad (n\to\infty). \quad (2.23)$$

但由本章定理 1.1,

$$\frac{1}{n}(\varepsilon(n))'(\varepsilon(n))\to\sigma^2 \text{ a. s.} \quad (n\to\infty).$$

从而最后证得

$$\frac{1}{n-P}(\varepsilon(n))'(I_n-X_nS_n^{-1}X_n')\varepsilon(n)\xrightarrow{P}\sigma^2. \quad (2.24)$$

定理证毕.

注意: 除去假设条件"存在 n_0, 使 $S_{n_0}>0$"外, 结果 $\hat{\sigma}_n^2\xrightarrow{P}\sigma^2$ 与设计矩阵 X_n 毫无关系. 如果用广义逆 S_n^- 来代替 S_n^{-1}, 则结论 $\hat{\sigma}_n^2\xrightarrow{P}\sigma^2$ 仍然成立, 此时就与 X_n 完全无关了.

关于线性模型中回归系数的 LS 估计的相合性、强相合性与极限分布问题, 近十多年来为许多作者所研究, 如 F. Eioker[4]、H. Drygas[5]、陈希孺[6]、Anderson And Taylor[7]、T. L. Lai、H. Robbins And C. Z. Wei[8] 以及陈桂景、黎子良和魏庆荣[9]等. 当 $\{e_i\}$ 为二阶矩一致有界的鞅差序列时, [8] 给出了 $\hat{\beta}_n\to\beta$(a. s.)的最弱的充分条件 $S_n^{-1}\to 0$, 并给出了 $\hat{\beta}_n$ 收敛的速度. [9] 在 [6]、[8] 等的基础上得出了更一般性的结果, 指出在 Gauss-Markov 模型下(第二章的 6.2), $\hat{\beta}_n\to\beta$(a. s.) 的 (本质上) 较好的充分条件是 $S_n^{-1}=O(\log^{-2}n(\log\log n)^{-(1+\varepsilon)})$, $\varepsilon>0$, 该文对 ε_i 的二阶矩非有限时也得出若干结果. 关于 σ^2 估计的强相合性的充要条件, 陈希孺[6]和赵林城[10]给出为

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\int_{|x|\geq\sqrt{n}}x^2\,dP_k=0, \quad (2.25)$$

$$\lim_{n\to\infty}\frac{1}{n^2}\sum_{k=1}^{n}\int_{|x|<\sqrt{n}}x^4\,dP_k=0, \quad (2.26)$$

对任意的 $\varepsilon>0$,

$$\sum_{k=0}^{\infty}P(|U_k^*|\geq\varepsilon)<\infty. \quad (2.27)$$

其中$U_k=\frac{1}{2^k}\sum_{i=2^{k+1}}^{2^{k+1}}(e_i^2-\sigma^2)$，$U_k^*$为$U_k$的对称化.

三、密度函数估计的相合性

由于客观实际错综复杂，使在统计判决问题中"假定总体服从某分布族"有时并不符合实际. 因此，有时需要利用观察数据来直接估计密度函数，以便得到总体的某些特征值的较好的估计. 无论从现代科学技术的应用需要上，还是从数理统计理论研究（如现代发展很快的自适应估计[11]、[12]，本书讨论的经验 Bayes 估计等）上，都需要密度函数的估计. 因此，近年来密度函数的估计受到很大的重视. 特别是六十年代初期 Parzen 提出核函数估计以来，密度函数的估计得到更大的发展，在这里，我们仅对 Parzen 的工作[13]作些扼要介绍。

设X_1, X_2, $\cdots$为$iid.$随机变量序列，具有密度函数$f(x)$（关于L测度），但不知其形式. 很早以前，人们就有用经验分布

$$F_n^*(x)=\frac{1}{n}(X_1,\ \cdots,\ X_n\text{中小于}x\text{的个数})$$

来估计理论分布$F(x)=\int_{-\infty}^{x}f(t)dt$的. 这样，容易想到利用经验分布来估计密度函数，采用如下形式

$$\begin{aligned}\hat{f}_n(x)&=[F_n^*(x+h)-F_n^*(x-h)]/2h\\&=\frac{1}{2h}\int_{x-h}^{x+h}dF_n^*(t)=\int_{-\infty}^{\infty}\frac{1}{h}K\left(\frac{x-t}{h}\right)dF_n^*(t)\\&=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K\left(\frac{x-x_i}{h}\right),\end{aligned}\tag{2.28}$$

其中

$$K\left(\frac{y}{h}\right)=\begin{cases}\frac{1}{2}, & \text{当}\left|\frac{y}{h}\right|\leqslant 1;\\0, & \text{当}\left|\frac{y}{h}\right|>1.\end{cases}\tag{2.29}$$

$K(\cdot)$称为估计的核. (2.29)是一种特殊的核函数. Parzen 就是从(2.28)式出发把核函数由(2.29)式的形式加以推广了. 除(2.29)形式的核以外，现在常见的核函数有如下几种:

核函数 $K(x)$	$g(t)=\int_{-\infty}^{\infty} e^{ity}K(y)dy$	$\frac{1}{2\pi}\int_{-\infty}^{\infty} g(t)dt$
$\begin{cases}1-\lvert x\rvert, & 当 \lvert x\rvert\leqslant 1;\\ 0, & 其它.\end{cases}$	$\left(\frac{\sin\frac{t}{2}}{\frac{t}{2}}\right)^2$	$\frac{2}{3}$
$\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$	$e^{-\frac{t^2}{2}}$	$\frac{1}{2\sqrt{\pi}}$
$\frac{1}{2}e^{-\lvert x\rvert}$	$\frac{1}{1+t^2}$	$\frac{1}{2}$
$\frac{1}{\pi}\frac{1}{1+x^2}$	$e^{-\lvert t\rvert}$	$\frac{1}{\pi}$
$\frac{1}{2}\left(\frac{\sin\frac{x}{2}}{\frac{x}{2}}\right)^2$	$\begin{cases}1-\lvert t\rvert, & 当 \lvert t\rvert\leqslant 1;\\ 0 & 其它.\end{cases}$	$\frac{1}{3\pi}$

为了使核函数能更好地估计密度函数，要求核函数及 h 满足如下条件：

(1) $K(x)\geqslant 0$ 且 $\int_{-\infty}^{\infty}K(x)dx=1$;

$$\sup_x K(x)<\infty, \int_{-\infty}^{\infty}K^2(x)dx<\infty; \tag{2.30}$$

$$\lim_{|x|\to\infty} K(x)x=0.$$

(2) $\lim\limits_{n\to\infty} h(n)=0,\quad \lim\limits_{n\to\infty} nh(n)=\infty.$

定理2.4 若 $X_1, X_2, \cdots$ 为 $iid.$ 随机序列，其分布具有有界的密度函数 $f(x)$，则利用满足条件(1)和(2)的核函数 $K(y)$ 给出的密度估计

$$\begin{aligned}\hat{f}_n(x)&=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{h(n)}K\left(\frac{x-x_i}{h(n)}\right)\\&=\frac{1}{h(n)}\int_{-\infty}^{\infty}K\left(\frac{x-t}{h(n)}\right)dF_n^*(t)\end{aligned} \tag{2.31}$$

是 $f(x)$ 的渐近无偏估计和相合估计.

证

$$\begin{aligned}
E\hat{f}_n(x) &= \frac{1}{h(n)} E\left\{K\left(\frac{x-X}{h(n)}\right)\right\} \\
&= \frac{1}{h(n)}\int_{-\infty}^{\infty} K\left(\frac{x-t}{h(x)}\right) f(t)\,dt \\
&= \frac{1}{h(n)}\int_{-\infty}^{\infty} K\left(\frac{y}{h(n)}\right) f(x-y)\,dy \quad (x-t=y) \\
&= \int_{-\infty}^{\infty} K(z) f(x-h(n)z)\,dz \quad \left(\frac{y}{h(n)}=z\right). \qquad (2.32)
\end{aligned}$$

对于任给的 $\varepsilon>0$, 由(1)可知, 存在充分大的 T_0, 使

$$\int_{|z|>T_0} K(z)\,dz \leqslant \varepsilon/4M,$$

此处 $M=\sup\limits_x f(x)$. 同时利用实变函数中的鲁金定理, 可证

$$\lim_{h\to 0}\int_{-T_0}^{T_0} K(z) f(x-hz)\,dz = f(x)\int_{-T_0}^{T_0} K(z)\,dz.$$

从而

$$\begin{aligned}
&|E\hat{f}_n(x)-f(x)| \\
&\leqslant \left|\int_{-T_0}^{T_0} K(z) f(x-h(n)z)\,dz - \int_{-T_0}^{T_0} K(z) f(x)\,dz\right| \\
&\quad +2\int_{|z|>T_0} K(z)\,dz\cdot M \\
&\leqslant \left|\int_{-T_0}^{T_0} K(z) f(x-h(n)z)\,dz - f(x)\int_{-T_0}^{T_0} K(z)\,dz\right| \\
&\quad +\frac{\varepsilon}{2} \to \frac{\varepsilon}{2} (n\to\infty). \qquad (2.33)
\end{aligned}$$

由于 ε 的任意性, 故 $\lim\limits_{n\to\infty} E\hat{f}_n(x)=f(x)$. 这便说明了 $\hat{f}_n(x)$ 为 $f(x)$ 的渐近无偏估计.

另一方面, 利用 $X_1, \cdots, X_n$ 的独立性, 有

$$\operatorname{Var}\hat{f}_n = \frac{1}{n}\frac{1}{h^2(n)}\left[EK^2\left(\frac{x-X}{h(n)}\right) - \left(EK\left(\frac{x-X}{h(n)}\right)\right)^2\right], \qquad (2.34)$$

用证明 $\hat{f}_n$ 为渐近无偏的同样办法, 可证

$$\lim_{n\to\infty}\frac{1}{h(n)} E\left[K^2\left(\frac{x-X}{h(n)}\right)\right] = f(x)\int_{-\infty}^{\infty} K^2(x)\,dx, \qquad (2.35)$$

由(2.33)～(2.35)，并注意条件(1)和(2)，得

$$\lim_{n\to\infty} E(\hat{f}_n-f(x))^2=\lim_{n\to\infty} \mathrm{Var}\,\hat{f}_n+\lim_{n\to\infty}(E\hat{f}_n-f(x))^2=0. \tag{2.36}$$

这说明对一切x，$\hat{f}_n$均方收敛于$f(x)$. 因此，$\hat{f}_n \xrightarrow{P} f(x)\ (n\to\infty)$. 定理证毕.

定理 2.5 假定定理2.4的条件成立，并进一步假设 i) $\lim_{n\to\infty} nh^2(n)=\infty$; ii) 核的 Fourier 变换 $g(t)$ 绝对可积; iii) $f(x)$ 一致连续. 则对任一 $\varepsilon>0$，有

$$\lim_{n\to\infty} P\{\sup_{-\infty<x<\infty} |\hat{f}_n(x)-f(x)|>\varepsilon\}=0. \tag{2.37}$$

证 记 $\varphi_n(t)=\int_{-\infty}^{\infty} e^{itx}\,dF_n^*(x)=\frac{1}{n}\sum_{j=1}^{n} e^{itx_j}$.

利用 Fourier 变换关于卷积定理知，$\hat{f}_n(x)$ 的 Fourier 变换是 $\varphi_n(t)g(th(n))$，再利用 Fourier 变换的反演定理，

$$\hat{f}_n(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\varphi_n(t)g(th(n))\,dt; \tag{2.38}$$

$$\begin{aligned}
&E\left[\sup_{-\infty<x<\infty} |\hat{f}_n(x)-E\hat{f}_n|\right]\\
&\leqslant \frac{1}{2\pi}\int_{-\infty}^{\infty} K(ht)E|\varphi_n(t)-E\varphi_n(t)|\,dt\\
&\leqslant \frac{1}{2\pi}\int_{-\infty}^{\infty} K(ht)\{E|\varphi_n(t)-E\varphi_n(t)|^2\}^{\frac{1}{2}}\,dt\\
&\leqslant \frac{1}{2\pi}\int_{-\infty}^{\infty} K(ht)\left\{\frac{1}{n}E(e^{itX}-Ee^{itX})(e^{-itX}-Ee^{-itX})\right\}^{\frac{1}{2}}dt\\
&\leqslant \frac{1}{2\pi n^{\frac{1}{2}}}\int_{-\infty}^{\infty} K(ht)\,dt\to 0 \quad (n\to\infty).
\end{aligned} \tag{2.39}$$

由(2.33)式，当$f(x)$为一致连续时，有

$$\sup_{-\infty<x<\infty} |E\hat{f}_n(x)-f(x)|\to 0 \quad (n\to\infty). \tag{2.40}$$

合并(2.39)与(2.40)得

$$E\left[\sup_{-\infty<x<\infty}|\hat{f}_n(x)-f(x)|\right]\leqslant E\left[\sup_{-\infty<x<\infty}|\hat{f}_n(x)-E\hat{f}_n(x)|\right]$$

$$+\sup_{-\infty<x<\infty}|E\hat{f}_n(x)-f(x)|\to 0\quad(n\to\infty).$$

从而可推得(2.37). 定理证毕.

注 如果$f(x)$有唯一的极大值点x_0, 我们取$\hat{f}_n(x)$的最大值点$\hat{x}_0(n)$, 则

$$|\hat{f}_n(\hat{x}_0(n))-f(x_0)|=|\sup_x\hat{f}_n(x)-\sup_x f(x)|$$

$$\leqslant\sup_x|\hat{f}_n(x)-f(x)|\xrightarrow{P}0\quad(n\to\infty).$$

故对任意的$\varepsilon>0$, 有$\lim\limits_{n\to\infty}P[|\hat{x}_0(n)-x_0|<\varepsilon]=1$. 这就给出了$f(x)$的极大值点的相合估计.

关于密度函数的估计, 特别是收敛速度的研究, 已有很多成果,有兴趣的读者可参考[14].

§3 估计的渐近有效性与最好渐近正态估计

Fisher 在 1925 年[15]就猜想渐近分布为正态的相合估计的渐近方差不小于C-R下界. 达到C-R下界的渐近正态估计, 称为**最好的渐近正态估计**, 记为 B. A. N 估计, 一般也称为**渐近有效估计**. 后来出现了一些超有效估计的例子. 这种估计是渐近正态的,但渐近方差不仅达到C-R下界,而且在有一些点上比C-R下界还小.

一、例子与定义

例3.1 设$X_1, \cdots, X_n, \cdots$为$iid.$随机序列, 服从正态分布$N(\theta, 1)$ $(-\infty<\theta<\infty)$. 我们已经知道$\overline{X}_n=\frac{1}{n}\sum\limits_{i=1}^n x_i$为$\theta$的完全充分统计量. 一直认为是$\theta$的优良估计. 现在考虑另一估计量

$$T_n=\begin{cases}\frac{1}{2}\overline{X}_n, & 若\ |\overline{X}_n|\leqslant\frac{\log n}{\sqrt{n}};\\ \overline{X}_n, & 其它.\end{cases}$$

当 $n>1$ 时,它是 θ 的有偏估计, $E_\theta T_n=\theta+b_n(\theta)$, 当 $\theta\neq 0$. 其中

$$b_n(\theta)=\frac{1}{2\sqrt{2\pi n}}\left(e^{-\frac{n\left(\frac{\log n}{\sqrt{n}}-\theta\right)^2}{2}}-e^{-\frac{n\left(\frac{\log n}{\sqrt{n}}+\theta\right)^2}{2}}\right)$$

$$-\frac{\theta}{2}\left[\Phi\left(\sqrt{n}\left(\frac{\log n}{\sqrt{n}}-\theta\right)\right)-\Phi\left(-\sqrt{n}\left(\frac{\log n}{\sqrt{n}}+\theta\right)\right)\right].$$

下面证明

$$P(\sqrt{n}\,(T_n-\theta)<x)\to\begin{cases}\Phi(2x), & 当\ \theta=0;\\ \Phi(x), & 当\ \theta\neq 0.\end{cases}$$

以后为方便起见,用 $\mathscr{L}(X_n)$ 表示 X_n 的分布函数. 若 X_n 的分布函数以分布函数 $F(x)$ 为极限分布,则记为 $\mathscr{L}(X_n)\to F(x)$. 用此记号,上式可表示为

$$\mathscr{L}(\sqrt{n}\,(T_n-\theta))\to\begin{cases}\Phi(2x), & 当\ \theta=0;\\ \Phi(x), & 当\ \theta\neq 0.\end{cases}$$

当 $\theta=0$ 时,

$$P_0(\sqrt{n}\ T_n<x)=P_0\left(\frac{\sqrt{n}}{2}\overline{X}_n<x,\ |\sqrt{n}\ \overline{X}_n|\leqslant\log n\right)$$

$$+P_0(\sqrt{n}\ \overline{X}_n<x,\ |\sqrt{n}\,\overline{X}_n|>\log n)\to\Phi(2x)\quad(n\to\infty).$$

这里主要用到 $P(|\sqrt{n}\ \overline{X}_n|>\log n)\to 0\quad(n\to\infty)$.

当 $\theta\neq 0$ 时, 由于当 $n\to\infty$ 时, $\frac{\log n}{\sqrt{n}}\to 0$, 有

$$P_\theta\left(|\overline{X}_n|>\frac{\log n}{\sqrt{n}}\right)\geqslant P_\theta\left(|\overline{X}_n-\theta|\leqslant\frac{1}{2}|\theta|\right)\to 1\ (n\to\infty),$$

故 $\lim\limits_{n\to\infty}P_\theta\left(|\overline{X}_n|>\frac{\log n}{\sqrt{n}}\right)=1$. 于是

$$P_\theta(\sqrt{n}\ (T_n-\theta)<x)=P_\theta\left(\sqrt{n}\left(\frac{1}{2}\overline{X}_n-\theta\right)<x;\ |\overline{X}_n|\leqslant\frac{\log n}{\sqrt{n}}\right)$$

$$+P_\theta\left(\sqrt{n}\,(\overline{X}_n-\theta)<x,\ |\overline{X}_n|>\frac{\log n}{\sqrt{n}}\right)\to\Phi(x)\quad(n\to\infty).$$

这就说明了, 当 $\theta=0$ 时, 渐近方差比 C-R 下界 1 还小, 当 $\theta\neq 0$ 时,刚好达到 C-R 下界.

这个例子说明了，只是限制在 $n^{\frac{1}{2}}$ 阶相合估计类中比较，还有缺陷，有必要对估计类进一步加以限制.

在本节中，记号 $\{\mathscr{X}_n, \mathscr{B}_{x_n}, \mathscr{P}_n\}$，$\mathscr{P}_n=\{P_{n,\theta}, \theta\in\Theta\}$，与 §1 相同. $\hat{\theta}_n$ 为依赖于 $x_n\in\mathscr{X}_n$ 的 θ 的估计量，$\{C_n\}$ 是一列非奇异的 p 阶方阵，与 θ 无关，且 $C_n^{-1}\to 0$ $(n\to\infty)$，C_n 的各特征根为正的.

定义 3.1 设 $\hat{\theta}_n$ 是 θ 的 C_n^{-1} 阶相合估计，$F_\theta(y)$ 是 p 维非退化的分布函数，$\theta\in\Theta$. 如果对一切的 $\theta\in\Theta$ 有

$$\lim_{n\to\infty} P_{n,\theta}\{C_n(\hat{\theta}_n-\theta)<y\}=F_\theta(y), \tag{3.1}$$

则称 $F_\theta(y)$ 为 $\hat{\theta}_n$ 的 C_n^{-1} 阶渐近分布，或简称渐近分布.

定义 3.2 设 $\hat{\theta}_n$ 为 C_n^{-1} 阶相合估计. 如果对一切 $\theta\in\Theta, a\in R^p$，$l\in R^p$，且 $\|l\|=1$，有

$$\lim_{n\to\infty} P_{n,\theta+C_n^{-1}a}\ (l'C_n(\hat{\theta}_n-\theta-C_n^{-1}a)<0)=\frac{1}{2}. \tag{3.2}$$

则称 $\hat{\theta}_n$ 为 θ 的渐近中位无偏估计.

从此定义出发，不难验证上面所举的超有效的正态分布均值估计的例子就不是渐近中位无偏的，具体验算，留给读者.

根据假设检验的 Neyman-Pearson 基本引理，若对零假设 $H_0: P_{n,\theta}$. 及对立假设 $P_{n,\theta_0+C_n^{-1}a}$ 进行检验，给定显著水平 α，就存在基本唯一的最大功效检验 ϕ_n. 根据对偶原则，在给定接受对立假设的概率时，存在基本唯一的检验，使拒绝零假设的概率达到最小. 因此，可以构造如下的检验 ϕ_n，使它满足

$$E_{n,\theta_0+C_n^{-1}a}\{\phi_n(X_n)\}\to\frac{1}{2}\quad(n\to\infty). \tag{3.3}$$

且使得 $E_{n,\theta_0}\{\phi_n(X_n)\}$ 达到最小.

现在取任意非零向量 $a\in R^p$ 以及实数 $t>0$，定义函数

$$\beta_{\theta_0}(t, a)=\sup_b \liminf_{n\to\infty} E_{n,\theta_0}\{\phi_n(X_n)\}, \tag{3.4}$$

其中 $\phi_n(x_n)$ 是对零假设 $H_0:\theta=\theta_0$，对立假设 $K:\theta=\theta_0+tC_n^{-1}b$ 的满足(3.3)式的最优检验，b 为满足 $a'b=1$ 的 p 维向量.

现考虑 θ 的任一 C_n^{-1} 阶渐近中位无偏的相合估计 $\hat{\theta}_n(x_n)$. 定

义拒绝 $\theta=\theta_0$ 的区域 $R_{n,a,t}=\{x_n: a'C_n(\hat{\theta}_n-\theta_0)<t\}$ (当 $t<0$ 时), $R_{n,a,t}=\{x_n: a'C_n(\hat{\theta}_n-\theta_0)>t\}$ (当 $t>0$). 故当 $t<0$ 时,

$$\begin{aligned}P_{n,\theta_0+tC_n^{-1}b}\{R_{n,a,t}\} &= P_{n,\theta_0+tC_n^{-1}b}\{a'C_n(\hat{\theta}_n-\theta_0-tC_n^{-1}b)<0\}\\&=\alpha_n(t);\\P_{n,\theta_0+tC_n^{-1}b}\{R_{n,a,t}\} &= P_{n,\theta_0+tC_n^{-1}b}\{a'C_n(\hat{\theta}_n-\theta_0-tC_n^{-1}b)>0\}\\&=\alpha_n(t) \quad (\text{当 } t>0).\end{aligned}$$

根据渐近中位无偏的性质, 有 $\lim\limits_{n\to\infty}\alpha_n(t)=\frac{1}{2}$. 记

$$G_{\theta_0}^{\hat{\theta}_n}(t,\ a)=\liminf_{n\to\infty} P_{n,\theta_0}(R_{n,a,t}).$$

由(3.4)中 $\beta_{\theta_0}(t,\ a)$ 的定义, 显然有

$$G_{\theta_0}^{\hat{\theta}_n}(t,\ a)\geqslant\beta_{\theta_0}(t,\ a). \tag{3.5}$$

定义 3.3 如果 θ 的一个 C_n^{-1} 阶渐近中位无偏相合估计 $\hat{\theta}_n$ 当 $n\to\infty$ 时满足条件 $G_{\theta_0}^{\hat{\theta}_n}(t,\ a)=\beta_{\theta_0}(t,\ a)$, 则称 $\hat{\theta}_n$ 为渐近有效估计(由[16]所给出的定义).

二、最好渐近正态估计(B. A. N. E)

一个 θ 的 $(\sqrt{n})^{-1}$ 阶相合估计 $\hat{\theta}_n$, 若使 $\mathscr{L}_\theta(\sqrt{n}(\hat{\theta}_n-\theta))$ 以 $N(0,\ A(\theta))$ 为极限分布, 其中 $A(\theta)=I^{-1}(\theta)$, 则称 $\hat{\theta}_n$ 为最好渐近正态估计. $I(\theta)$ 为 Fisher 信息矩阵. 并假设 $g(\theta)$ 是 θ 的连续可微的 k 维向量函数($k\leqslant p$). $\hat{g}_n(x_n)$ 是 $g(\theta)$ 的 $(\sqrt{n})^{-1}$ 阶相合估计, 若 $\mathscr{L}_\theta(\sqrt{n}(\hat{g}_n-g(\theta)))$ 以 $N(0,\ B(\theta))$ 为其极限分布, 其中 $B(\theta)=\left(\frac{\partial g}{\partial\theta}\right)'I^{-1}(\theta)\left(\frac{\partial g}{\partial\theta}\right)$, 则称 $\hat{g}_n$ 为 $g(\theta)$ 的最好渐近正态估计.

下面我们采用竹内启[16]的思想来证明 θ 的 B. A. N. E 在一定条件下是渐近有效估计.

定理 3.1 设 $X_1,\ X_2,\ \cdots$ 为 $iid.$ 随机序列, X_1 具有分布 $P_\theta,\ \theta\in\Theta$. 对每一个 $\theta\in\Theta$, P_θ 具有关于 σ-有限测度 μ 的概率密度为 $f(x,\ \theta)=dP_\theta(x)/d\mu$. $f(x,\ \theta)$ 满足下列条件:

i) $\{f(x,\ \theta),\ \theta\in\Theta\}$ 具有共同支撑. 不妨设对一切的 $x\in\mathscr{X}$,

$f(x, \theta)>0$. 当 $\theta_1\neq\theta_2(\in\Theta)$时, $\mu(f(x, \theta_1)\neq f(x, \theta_2))>0$.

ii) $f(x, \theta)$ 在对每一个 x 固定时, 关于 θ 有一阶连续偏微分*).

iii) 对每个 $\theta_0\in\Theta$, 存在 θ_0 的邻域 $V_{\theta_0}\subset\Theta$, 使得

$$\left|\frac{f(x, \theta)}{f(x, \theta_0)}-1\right|\leqslant A(x, \theta_0)\,|\theta-\theta_0| \quad (\text{当 } \theta\in V_{\theta_0} \text{ 时});$$

$$|Z(x, \theta)|\triangleq\left|\frac{\partial}{\partial\theta}\log f(x, \theta)\right|\leqslant B(x, \theta_0) \quad (\text{当 } \theta\in V_{\theta_0} \text{ 时}).$$

此处 A、B 满足条件 $E_{\theta_0}A(x, \theta_0)<\infty$, $E_{\theta_0}B^2(x,\theta_0)<\infty$, $E_{\theta_0}A(x, \theta_0)B^2(x, \theta_0)<\infty$.

那么, θ 的具有渐近中位无偏性的 B. A. N. E 是渐近有效估计.

证 根据 $\sqrt{n}^{-1}$ 阶渐近有效估计的定义, 重要的是找出 $\beta_{\theta_0}(t, a)$ 的表达式以及 $\beta_{\theta_0}(t, a)$ 与 $I(\theta_0)$ 的联系. 考虑建立 H_0: $\theta=\theta_0$, $K:\theta=\theta_0+t(\sqrt{n})^{-1}b$ 的最优检验. 令

$$W_n=\sum_{i=1}^{n}\log\frac{f(x, \theta_0+t(\sqrt{n})^{-1}b)}{f(x, \theta_0)}. \tag{3.6}$$

那么由 Neyman-Pearson 基本引理知, 最优检验的拒绝区域形式为 $W_n>C_n$. 首先的问题是能否找到 C_n, 使得对任意的 $t\in R^1$ 与 $b\in R^p$, 有

$$P_{\theta_0+t(\sqrt{n})^{-1}b}(W_n>C_n)\to\frac{1}{2} \quad (n\to\infty). \tag{3.7}$$

根据引理 1.1, $I(\theta_0+tn^{-\frac{1}{2}}b)=I(\theta_0)+o(1)$,

$$E_{\theta_0+tn^{-\frac{1}{2}}b}\left\{\log\frac{f(x, \theta_0+tn^{-\frac{1}{2}}b)}{f(x, \theta_0)}\right\}=\frac{1}{2n}t^2b'I(\theta_0)b+o\left(\frac{1}{n}\right);$$

$$E_{\theta_0+tn^{-\frac{1}{2}}b}\left\{\log\frac{f(x, \theta_0+tn^{-\frac{1}{2}}b)}{f(x, \theta_0)}\right\}^2=\frac{1}{n}t^2b'I(\theta_0)b+o\left(\frac{1}{n}\right).$$

根据中心极限定理(见 Гнеденко «概率论教程»第九章; 并注意假设 iii))得

---

*) 此条件可减弱为: 对 a. s. P_θ, $\theta\in\Theta$ 的 x 成立.

$$P_{\theta_0+tn^{-\frac{1}{2}}b}\left(\frac{W_n-\frac{1}{2}t^2b'I(\theta_0)b}{|t|(b'I(\theta_0)b)^{\frac{1}{2}}}>0\right)\to\frac{1}{2}\quad(n\to\infty).\quad(3.8)$$

故取 $C_n=\frac{1}{2}t^2b'I(\theta_0)b$ 合乎所求.

另一方面,由引理 1.1 知,

$$E_{\theta_0}\left\{\log\frac{f(x,\ \theta_0+tn^{-\frac{1}{2}}b)}{f(x,\ \theta_0)}\right\}=-\frac{1}{2n}t^2b'I(\theta_0)b+o\left(\frac{1}{n}\right);\quad(3.9)$$

$$E_{\theta_0}\left\{\log\frac{f(x,\ \theta_0+tn^{-\frac{1}{2}}b)}{f(x,\ \theta_0)}\right\}^2=\frac{1}{n}t^2b'I(\theta_0)b+o\left(\frac{1}{n}\right).\quad(3.10)$$

由中心极限定理得

$$\begin{aligned}&P_{\theta_0}\left\{W_n>\frac{1}{2}t^2b'I(\theta_0)b\right\}\\&=P_{\theta_0}\left\{\frac{W_n+\frac{1}{2}t^2b'I(\theta_0)b}{|t|(b'I(\theta_0)b)^{\frac{1}{2}}}>|t|(b'I(\theta_0)b)^{\frac{1}{2}}\right\}\\&\to\Phi(-|t|(b'I(\theta_0)b)^{\frac{1}{2}})\quad(\text{当 } n\to\infty \text{ 时}).\quad(3.11)\end{aligned}$$

在 $a'b=1$ 的限制下，使 $b'I(\theta_0)b$ 达到最小的 b 为 $I^{-1}(\theta_0)a/a'I^{-1}(\theta_0)a$. 故

$$\beta_{\theta_0}(t,\ a)=\Phi(-|t|/(a'I^{-1}(\theta_0)a)^{\frac{1}{2}}).\quad(3.12)$$

由假设, $\hat{\theta}_n$ 具有渐近中位无偏性,并且

$$\mathscr{L}_{\theta_0}(\sqrt{n}(\hat{\theta}_n-\theta_0))\to N(0,\ I^{-1}(\theta_0))\quad(\text{当 } n\to\infty \text{ 时}).$$

故有

$$\lim_{n\to\infty}P_{\theta_0}(R_{n,a,t})=\Phi(-|t|/(a'I^{-1}(\theta_0)a)^{\frac{1}{2}})=G_{\theta_0}^{\hat{\theta}_n}(t,\ a).\quad(3.13)$$

此处

$$R_{n,a,t}=\begin{cases}((x_1,\ \cdots,\ x_n)':\ a'\sqrt{n}(\hat{\theta}_n-\theta_0)>t)&(\text{当 } t>0)\\((x_1,\ \cdots,\ x_n)':\ a'\sqrt{n}(\hat{\theta}_n-\theta_0)<t)&(\text{当 } t<0).\end{cases}$$

故 $\beta_{\theta_0}(t, a)=G_{\theta_0}^{\hat{\theta}_n}(t, a)$ 对所有的 $\theta_0\in\Theta$, $t\in R$, $a\neq 0(\in R^p)$ 成立.这就证明了 $\hat{\theta}_n$ 是 θ 的渐近有效估计.

系 3.1 在定理 3.1 条件下，对 θ 的任意的渐近中位无偏估计 $\tilde{\theta}_n$, 若 $\mathscr{L}(\sqrt{n}(\tilde{\theta}_n-\theta))\to N(0, A(\theta))$，则 $A(\theta)\geqslant I^{-1}(\theta)$ 对一切 $\theta\in\Theta$ 成立.

证 由定理 3.1, $\beta_{\theta_0}(t, a)=\Phi(-|t|/(a'I(\theta_0)a)^{\frac{1}{2}})$. 不难看到

$$G_{\theta_0}^{\tilde{\theta}_n}(t, a)=\Phi(-|t|/(a'A(\theta_0)a)^{\frac{1}{2}})\geqslant\beta_{\theta_0}(t, a)$$
$$=\Phi(-|t|/(a'I^{-1}(\theta_0)a)^{\frac{1}{2}})$$

对一切 $t\in R'$, $a(\neq 0)\in R^p$, $\theta_0\in\Theta$ 成立. 故

$$(a'A(\theta_0)a)^{\frac{1}{2}}\geqslant(a'I^{-1}(\theta_0)a)^{\frac{1}{2}} \quad (a(\neq 0)\in R^p,\ \theta_0\in\Theta).$$

从而 $A(\theta_0)\geqslant I^{-1}(\theta_0)$, $\theta_0\in\Theta$. 此系证毕.

这个系说明了,为什么把极限分布为 $N(0, I^{-1}(\theta))$ 的估计叫做 B. A. N 估计.

下面一个定理为我们提供了一个把一般估计改进为 B. A. N 估计的办法.

定理 3.2 假定定理 3.1 中关于分布族的假设条件 i)~iii) 成立. 进一步假设 iv) $\dfrac{\partial Z(x, \theta)}{\partial\theta}$ 对每一个 $x\in\mathscr{X}$ 及 $\theta\in\Theta$ 存在且有限,而且对每一个 $\theta_0\in\Theta$ 存在邻域 $V_{\theta_0}\subset\Theta$ 和关于 P_{θ_0} 可积函数 $H(x, \theta_0)$, 使

$$\left|\frac{\partial Z(x, \theta)}{\partial\theta}-\frac{\partial Z(x, \theta_0)}{\partial\theta_0}\right|\leqslant H(x, \theta_0)|\theta-\theta_0|. \tag{3.14}$$

v) 对一切 $\theta\in\Theta$, 有

$$\frac{\partial^2}{\partial\theta^2}\int f(x, \theta)d\mu=\int\frac{\partial^2}{\partial\theta^2}f(x, \theta)d\mu. \tag{3.15}$$

这里 $\dfrac{\partial^2}{\partial\theta^2}$ 表示微分算子 $\left(\dfrac{\partial^2}{\partial\theta_i\partial\theta_j}\right)_{P\times P}$.

若 $\tilde{\theta}_n$ 是 θ 的 $n^\alpha I_P\left(\alpha>\dfrac{1}{4}\right)$ 级相合估计,则当 n 充分大时,下列估计 $\hat{\theta}_n$ 是 B. A. N 估计:

$$\hat{\theta}_n=\tilde{\theta}_n-\frac{1}{\sqrt{n}}\hat{B}_n^{-1}(\tilde{\theta}_n)\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}Z(x_i,\ \tilde{\theta}_n)\right). \tag{3.16}$$

此处 $\hat{B}_n(\tilde{\theta}_n)=\frac{1}{n}\sum_{i=1}^{n}\left.\frac{\partial Z(x_i,\ \theta)}{\partial\theta}\right|_{\theta=\tilde{\theta}_n}$.

注意 $\hat{B}_n(\tilde{\theta}_n)$的逆阵可能不存在，此时在$\hat{\theta}_n$的表示式中可把$\hat{B}_n^{-1}$改为广义逆. 不过，下面我们要证明 $\hat{B}_n(\tilde{\theta}_n)\xrightarrow{P}-I(\theta)$，而$I(\theta)>0$. 从而可知，当$n$很大时，使$\hat{B}^{-1}(\tilde{\theta}_n)$存在的概率可任意接近1.

证 根据假设 v)：

$$0=\frac{\partial^2}{\partial\theta^2}\int_{\mathscr{X}}f(x,\ \theta)d\mu=\int_{\mathscr{X}}\frac{\partial^2}{\partial\theta^2}f(x,\ \theta)d\mu$$

$$=\int_{\mathscr{X}}\left[\frac{\partial^2}{\partial\theta^2}\log f(x,\ \theta)\right.$$

$$\left.+\left(\frac{\partial\log f(x,\ \theta)}{\partial\theta}\right)\left(\frac{\partial\log f(x,\ \theta)}{\partial\theta}\right)'\right]f(x,\ \theta)d\mu$$

$$=\int_{\mathscr{X}}\frac{\partial Z(x,\ \theta)}{\partial\theta}f(x,\ \theta)+I(\theta).$$

故

$$E_\theta\left(\frac{\partial Z(X,\ \theta)}{\partial\theta}\right)=-I(\theta). \tag{3.17}$$

由于$\tilde{\theta}_n$是相合估计，以及假设 iv)$I(\theta)$ 连续，根据大数法则，当$n\to\infty$时，有

$$\begin{aligned}\hat{B}_n(\tilde{\theta}_n)&=\hat{B}_n(\theta_0)+O_P(|\tilde{\theta}_n-\theta_0|)\\&=-I(\theta_0)+O_P(|\tilde{\theta}_n-\theta_0|)+o_P(1);\end{aligned} \tag{3.18}$$

$$Z(x,\ \tilde{\theta}_n)=Z(x,\ \theta_0)+\frac{\partial Z(x,\ \theta_0)}{\partial\theta_0}(\tilde{\theta}_n-\theta_0)+o_P(|\tilde{\theta}_n-\theta_0|). \tag{3.19}$$

故根据$\hat{\theta}_n$的定义，有

$$\begin{aligned}\sqrt{n}\ (\hat{\theta}_n-\theta_0)&=\sqrt{n}\ (\tilde{\theta}_n-\theta_0)-\hat{B}_n^{-1}(\tilde{\theta}_n)\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}Z(x_i,\ \tilde{\theta}_n)\right)\\&=\sqrt{n}\ (\tilde{\theta}_n-\theta_0)-[\hat{B}_n^{-1}(\theta_0)+O_P(|\tilde{\theta}_n-\theta_0|)].\end{aligned}$$

$$\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}Z(x_i,\theta_0)+\sqrt{n}\hat{B}_n(\theta_0)(\tilde{\theta}_n-\theta_0)+\sqrt{n}\theta_P(|\tilde{\theta}_n-\theta_0|^2)\right]$$

$$=I^{-1}(\theta_0)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}Z(x_i,\ \theta_0)+o_P(1). \qquad (3.20)$$

因为

$$E_{\theta_0}Z(x_i,\ \theta_0)=0,\quad E_{\theta_0}\{(Z(x_i,\ \theta_0))(Z(x_i,\ \theta_0))'\}=I(\theta_0).$$

故 $\mathscr{L}_{\theta_0}(\sqrt{n}(\hat{\theta}_n-\theta_0))\to N(0,\ I^{-1}(\theta_0))$. 最后一步用到了定理 1.3.

三、非共同支撑分布族的渐近有效估计的例子

设 $X_1,\ X_2,\ \cdots$ 为 *iid*. 随机变量序列, X_1 具有非共同支撑的分布: $dP_\theta(x)=f(x,\ \theta)d\mu,\ \theta\in\Theta$. 例如 $P_\theta=R(0,\ \theta)$ $(0<\theta<\infty)$ 或 $dP_\theta(x)=\frac{1}{\beta}e^{-\frac{x-\theta}{\beta}}dx$ $(0<\beta<\infty,\ -\infty<\theta<\infty,\ x>\theta)$ 等等. 下面我们就个别例子来讨论非共同支撑分布族中的渐近有效估计问题.

定理 3.3 若 $r(x)>0$ 为 $[0,\ \infty)$ 上定义的连续函数, 考虑如下分布族

$$dP_\theta(x)=f(x,\ \theta)dx=\begin{cases}q(\theta)r(x)dx, & \text{当 } 0<x\leqslant\theta;\\ 0, & \text{其它}\end{cases}$$

$$(0<\theta<\infty).$$

$X_1,\ X_2,\ \cdots$ 为具有密度 $f(x,\ \theta)$ 的 *iid*. 随机序列. 则有

$$\beta_\theta(t,\ 1)=\begin{cases}1-\frac{1}{2}e^{r(\theta)q(\theta)t}, & \text{当 } 0<t<(r(\theta)q(\theta))^{-1}\log 2;\\ 0, & \text{当 } t\geqslant(r(\theta)q(\theta))^{-1}\log 2.\end{cases}$$

$$\beta_\theta(t,\ -1)=\frac{1}{2}e^{-r(\theta)q(\theta)t},\quad \text{当 } t>0. \qquad (3.21)$$

又若 $\hat{\theta}_n$ 为 θ 的 50% 置信上界, 则它是渐近有效估计.

证 因 $\prod_{i=1}^{n}f(x_i,\ \theta)=q^n(\theta)\prod_{i=1}^{n}r(x_i)I_{[\max_{1\leqslant i\leqslant n}x_i<\theta]}(x_1,\ \cdots,\ x_n)$

故 $x^*_{(n)}=\max_{1\leqslant i\leqslant n}x_i$ 为充分统计量. 根据 Neyman-Pearson 基本引理, 关于零假设 $H_0:\theta=\theta_0$, 对立假设 $K:\theta=\theta_0+n^{-1}t$ 的最优检验形

式为

$$\varphi(x^*_{(n)})=\begin{cases}1, & 当\ x^*_{(n)}>C_n,\ 对\ t>0;\ 当\ x^*_{(n)}<C_n,\ 对\ t<0;\\ 0, & 其它.\end{cases}$$

先考虑 $t>0$ 的情况: $X^*_{(n)}$ 的分布为

$$\left(\int_0^x f(y,\ \theta)dy\right)^n=P^n_\theta(x).$$

要使 $$P_{\theta_0+n^{-1}t}(X^*_{(n)}>C_n)=1-P^n_{\theta_0+n^{-1}t}(C_n)=\frac{1}{2},$$

从而可解得: 当 $C_n<\theta_0$ 时,

$$q^{-1}(C_n)=\int_0^{C_n} r(y)dy=q^{-1}(\theta_0+n^{-1}t)\left(\frac{1}{2}\right)^{\frac{1}{n}};$$

当 $0<t\leqslant(r(\theta_0)q(\theta_0))^{-1}\log 2$,

$$\begin{aligned}P_{\theta_0}(X^*_{(n)}>C_n)&=1-\left(q(\theta_0)q^{-1}(\theta_0+n^{-1}t)\left(\frac{1}{2}\right)^{\frac{1}{n}}\right)^n\\&=1-\frac{1}{2}\left(1+\int_{\theta_0}^{\theta_0+n^{-1}t} r(y)dy\Big/\int_0^{\theta_0} r(y)dy\right)^n\\&=1-\frac{1}{2}\left(1+\frac{t}{n}\,r(\theta_0)q(\theta_0)+o\left(\frac{1}{n}\right)\right)^n\\&\to 1-\frac{1}{2}\,e^{r(\theta_0)q(\theta_0)t}.\end{aligned}\tag{3.22}$$

在 $C_n>\theta_0$ 时, 或 $t\geqslant(r(\theta_0)q(\theta_0))^{-1}\log 2$ 时, $P_{\theta_0}(X^*_{(n)}>C_n)\to 0$. 那么, 由(3.22)就有 $\beta_{\theta_0}(t,\ 1)=1-\frac{1}{2}\,e^{r(\theta_0)q(\theta_0)t}$.

当 $t<0$ 时,

$$\begin{aligned}P_{\theta_0}(X^*_n<C_n)&=\left(q(\theta_0)q^{-1}(\theta_0+n^{-1}t)\left(\frac{1}{2}\right)^{\frac{1}{n}}\right)^n\\&=\frac{1}{2}\left(1-\int_{\theta_0+n^{-1}t}^{\theta_0} r(y)dy\Big/\int_0^{\theta_0} r(y)dy\right)^n\\&\to\frac{1}{2}\,e^{-r(\theta_0)q(\theta_0)t}\quad(n\to\infty).\end{aligned}\tag{3.23}$$

由此式得到

$$\beta_{\theta_0}(t,\ -1)=\frac{1}{2}\,e^{-r(\theta_0)q(\theta_0)t}.\tag{3.24}$$

现在采用满足方程

$$P_\theta^n(x^*_{(n)})=(q(\theta)q^{-1}(x^*_{(n)}))^n=\frac{1}{2}$$

的估计 $\hat{\theta}_n$，也即 $q(\hat{\theta}_n)=\left(\frac{1}{2}\right)^{\frac{1}{n}}q(x^*_{(n)})$。因为 $q(\theta)$ 为严格单调下降函数，故其解存在且唯一.

$$\begin{aligned}q^{-1}(\hat{\theta}_n)-q^{-1}(\theta)&=2^{\frac{1}{n}}q^{-1}(x^*_{(n)})-q^{-1}(\theta)\\&=r(\theta)(x^*_{(n)}-\theta)+\frac{\log 2}{n}q^{-1}(x^*_{(n)})\\&\quad+o\left(\frac{1}{n}\right)+o_P(x^*_{(n)}-\theta)\end{aligned}$$

$\left(\text{因为 } q^{-1}(x^*_{(n)})-q^{-1}(\theta)=r(\theta)(x^*_{(n)}-\theta)+o_P(x^*_{(n)}-\theta),\ -1+2^{\frac{1}{n}}=\frac{\log 2}{n}+o\left(\frac{1}{n}\right)\right)$. 另一方面,

$$q^{-1}(\hat{\theta}_n)-q^{-1}(\theta)=r(\theta)(\hat{\theta}_n-\theta)+o_P(\hat{\theta}_n-\theta),$$

所以 $n(\hat{\theta}_n-\theta)=n(x^*_{(n)}-\theta)+(\log 2)q^{-1}(\theta)r^{-1}(\theta)+o_P(1)$.

因此，$\mathscr{L}(n(\hat{\theta}_n-\theta))$ 与 $\mathscr{L}(n(x^*_{(n)}-\theta)+(\log 2)q^{-1}(\theta)r^{-1}(\theta))$ 具有相同的极限分布. 但对于 $t<0$,

$$\begin{aligned}P_\theta\{n(x^*_{(n)}-\theta)<t\}&=P_\theta\left(x^*_{(n)}<\theta+\frac{t}{n}\right)\\&=\left(q(\theta)q^{-1}\left(\theta+\frac{t}{n}\right)\right)^n\\&=\left(1+\frac{q^{-1}\left(\theta+\frac{t}{n}\right)-q^{-1}(\theta)}{q^{-1}(\theta)}\right)^n\\&\to e^{r(\theta)q(\theta)t}\quad(n\to\infty).\end{aligned}$$

故当 $t\leqslant(r(\theta)q(\theta))^{-1}\log 2$ 时,

$$P_\theta(n(\hat{\theta}_n-\theta)<t)\to\frac{1}{2}e^{r(\theta)q(\theta)t}\quad(n\to\infty).$$

由 $G_\theta^{\hat{\theta}_n}(-1,t)$ 的定义，在 $t>0$ 时有

$$G_\theta^{\hat{\theta}_n}(-1,t)=\frac{1}{2}e^{-r(\theta)q(\theta)t}.$$

再由 $G_\theta^{\hat{\theta}_n}(1,t)$ 的定义，当 $0<t\leqslant(r(\theta)q(\theta))^{-1}\log 2$ 时，值为

$1-\frac{1}{2}e^{-r(\theta)q(\theta)t}$，其它情况为0. 这就证得

$$G_\theta^{\hat{\theta}_n}(1,\ t)=\beta_\theta(t,\ 1),\ G_\theta^{\hat{\theta}_n}(-1,\ t)=\beta_\theta(t,\ -1) \quad (3.25)$$

对一切 $t>0$ 及 $\theta\in\Theta$ 成立. 至于 $\hat{\theta}_n$ 为 θ 的渐近中位无偏估计，则是显然的. 从而 $\hat{\theta}_n$ 为 θ 的渐近有效估计. 定理证毕.

关于非共同支撑渐近有效估计的其它例子，可参考竹内启的著作[16].

四、其它 C_n 阶渐近有效估计——回归模型

我们考虑回归模型

$$y_i=x_i'\beta+\varepsilon_i \quad (i=1,\ 2,\ \cdots),$$

此处回归系数 $\beta\in R^P$ 未知，$x_i=(x_{i1},\ \cdots,\ x_{iP})'$ $(i=1,\ 2,\ \cdots)$为已知向量序列. 设存在 n_0，使 $n\geqslant n_0$ 时，$S_n=\sum_{i=1}^n x_ix_i'>0$. 并假定 $|x_i|\leqslant c<\infty$ $(i=1,\ 2,\ \cdots)$，$S_n^{-1}\to 0$ $(n\to\infty)$.

$\varepsilon_1,\ \varepsilon_2,\ \cdots$ 为 *iid*. 随机误差序列，具有密度 $f(x)$，它满足如下条件：

i) $f(x)>0$，$x\in R^1$，除个别 x 外，$f(x)$ 具有一阶连续导数.

ii) 对任意的 $x\in R^1$，皆存在适当小的 $\varDelta>0$，使

$$\left|\frac{f(x\pm\varDelta)}{f(x)}-1\right|\leqslant A(x)\varDelta;$$

$$|Z(x\pm\varDelta)|\triangleq\left|\frac{\partial}{\partial x}\log f(x\pm\varDelta)\right|\leqslant B(x).$$

其中 $$\int A(x)f(x)dx<\infty,\quad \int B^2(x)f(x)dx<\infty,$$

$$\int B^2(x)A(x)f(x)dx<\infty.$$

iii) $\int xf(x)dx=0$.

定理3.4 在上述条件下，若 β 的 $S_n^{\frac{1}{2}}$ 阶渐近中位无偏估计 $\hat{\beta}_n$ 有 $\mathscr{L}(S_n^{\frac{1}{2}}(\hat{\beta}_n-\beta))\to N(0,\ d^{-1}I_P)$ $(n\to\infty)$，则它是渐近有效估计，也是最好渐近正态估计. 此处

$$d=\int(f'(x))^2/f(x)dx,\quad S_n^{\frac{1}{2}}S_n^{\frac{1}{2}}=S_n.$$

证　首先求找 $\beta_\beta(a, t)$，记

$$Z_n=\sum_{i=1}^{n}\log\frac{f(y_i-x_i'(\beta+\Delta\beta))}{f(y_i-x_i'\beta)},\tag{3.26}$$

由引理 1.1,

$$E_{\beta+\Delta\beta}\left\{\log\frac{f(y_i-x_i'(\beta+\Delta\beta))}{f(y_i-x_i'\beta)}\right\}$$

$$=\frac{1}{2}d(\Delta\beta)'x_ix_i'(\Delta\beta)(1+o(1));$$

$$E_{\beta+\Delta\beta}\left\{\log\frac{f(y_i-x_i'(\beta+\Delta\beta))}{f(y_i-x_i'\beta)}\right\}^2$$

$$=d(\Delta\beta)'x_ix_i'(\Delta\beta)(1+o(1)).$$

所以

$$\mathrm{Var}_{\beta+\Delta\beta}\left\{\log\frac{f(y_i-x_i'(\beta+\Delta\beta))}{f(y_i-x_i'\beta)}\right\}$$

$$=d(\Delta\beta)'x_ix_i'(\Delta\beta)(1+o(1)).$$

注意此处 dx_ix_i' 刚好是 $f(y_i-x_i'\beta)$ 的信息阵．因此有

$$E_{\beta+\Delta\beta}\{Z_n\}=\frac{1}{2}d(\Delta\beta)'S_n(\Delta\beta)(1+o(1));\tag{3.27}$$

$$\mathrm{Var}_{\beta+\Delta\beta}\{Z_n\}=d(\Delta\beta)'S_n(\Delta\beta)(1+o(1)).\tag{3.28}$$

因为 $S_n^{-1}\to 0$, $|x_i|\leqslant C<\infty$ $(i=1, 2, \cdots)$，注意

$$E_{\beta+tS_n^{-\frac{1}{2}}b}\{Z_n\}=dt^2\|b\|^2/2(1+o(1));$$

$$\mathrm{Var}_{\beta+tS_n^{-\frac{1}{2}}b}\{Z_n\}=dt^2\|b\|^2(1+o(1)),$$

利用假设条件 iii）和中心极限定理，有

$$\mathscr{L}_{\beta+tS_n^{-\frac{1}{2}}b}\left(\left(Z_n-\frac{dt^2}{2}\|b\|^2\right)\right)\to N(0,\ dt^2\|b\|^2).\tag{3.29}$$

因此

$$\lim_{n\to\infty}P_{\beta+tS_n^{-\frac{1}{2}}b}\left(Z_n-\frac{1}{2}dt^2\|b\|^2\geqslant 0\right)=\frac{1}{2}.\tag{3.30}$$

同理有

$$E_\beta\{Z_n\}=-\frac{1}{2}dt^2\|b\|^2(1+o(1));\tag{3.31}$$

$$\mathrm{Var}_\beta\{Z_n\}=dt^2\|b\|^2(1+o(1)). \tag{3.32}$$

故有

$$\lim_{n\to\infty}P_\beta\left(Z_n-\frac{1}{2}dt^2\|b\|^2\geqslant 0\right)$$
$$=\lim_{n\to\infty}P_\beta\left\{Z_n+\frac{1}{2}dt^2\|b\|^2\geqslant dt^2\|b\|^2\right\}$$
$$=\Phi\left(\frac{-dt^2\|b\|^2}{d^{\frac{1}{2}}|t|\,\|b\|}\right)=\Phi\left(-d^{\frac{1}{2}}|t|\,\|b\|\right). \tag{3.33}$$

对任意的 $a\neq 0$, $a\in R^p$, 取 $b=\|a\|^{-2}a$, 可在限制条件 $a'b=1$ 使 $\Phi(-d^{\frac{1}{2}}|t|\,\|b\|)$ 达到最大,从而求得

$$\beta_\beta(t,\ a)=\Phi(-d^{\frac{1}{2}}|t|\,\|a\|^{-1}). \tag{3.34}$$

由假设条件 $\mathscr{L}_\beta(S_n^{\frac{1}{2}}(\hat{\beta}_n-\beta))\to N(0,\ d^{-1}I_P)$, 故

$$\lim_{n\to\infty}P_\beta\{a'S_n^{\frac{1}{2}}(\hat{\beta}_n-\beta)>|t|\}=\lim_{n\to\infty}P_\beta\{a'S_n^{\frac{1}{2}}(\hat{\beta}_n-\beta)<-|t|\}$$
$$=\Phi(-|t|d^{\frac{1}{2}}\|a\|^{-1})=\beta_\beta(t,\ a). \tag{3.35}$$

事实上,上式就是

$G_\beta^{\hat{\beta}_n}(t,\ a)=\beta_\beta(t,\ a)$, 对任意的 β, $a\in R^p$, $a\neq 0$, $t\in R^1$. 这就证明了 $\hat{\beta}_n$ 是渐近有效估计.

对其它任一 β 的估计 $\tilde{\beta}_n$, 若满足

$$\mathscr{L}(S_n^{\frac{1}{2}}(\tilde{\beta}_n-\beta))\to N(0,\ A(\beta)),$$

根据 Neyman-Pearson 基本引理,对任意的 $a\neq 0, \beta\in R^p$, $t\in R^1$, 有

$$G_\beta^{\tilde{\beta}_n}(t,\ a)\geqslant\beta_\beta(t,\ a),\ a'A(\beta)a\geqslant d^{-1}\|a\|^2.$$

也即有 $A(\beta)\geqslant d^{-1}I_P$. 这说明了 $\hat{\beta}_n$ 是最优渐近正态估计. 至此定理证毕.

至于 $\hat{\beta}_n$ 采用什么形式可达到 β 的渐近有效估计,用类似定理 3.2 的办法对最小二乘估计进行改造即可得.

§4* Bahadur 渐近有效性

Bahadur 在 1960 年[17] 提出了一种渐近有效性的概念，这种有效性是

比较估计与真值的差值落在 0 的邻域内的概率，当样本大小 n 无限增大以及邻域无限缩小时趋于 1 的速度. 本节介绍 Bahadur 这一渐近有效性的概念. 首先叙述分布族有共同支撑、Fisher 信息矩阵存在有限的情况，然后简述定义的扩充情况.

设 $X_1, X_2, \cdots$ 为从总体$(\mathscr{X}, \mathscr{B}_x, P_\theta, \theta\in\Theta)$ 中抽得的 $iid.$ 随机序列. Θ 是在 R^p 中的开区域. P_θ 关于 σ-有限测度 μ 有概率密度 $f(x, \theta)=dP_\theta/d\mu$ $(\theta\in\Theta)$. 我们假定: i) $f(x,\theta)$、$\theta\in\Theta$ 有共同支撑; ii) $Z(x,\theta)=\dfrac{\partial\log f(x,\theta)}{\partial\theta}$ 存在且有限; iii) $I(\theta)=E_\theta\{(Z(X, \theta))(Z(x, \theta))'\}>0$, 有限. 令 $\hat{\theta}_n$ 是 θ 的依赖于 $x_1, \cdots, x_n$ 的估计.

定义 4.1 若 $\hat{\theta}_n$ 是 θ 的相合估计，又若满足

$$\lim_{\varepsilon\to0}\sup\lim_{n\to\infty}\sup\frac{1}{n\varepsilon^2}\log P_\theta(\|\hat{\theta}_n-\theta\|\geqslant\varepsilon)\leqslant-\frac{1}{2}\frac{1}{\|I^{-1}(\theta)\|}\qquad(4.1)$$

对一切 $\theta\in\Theta$ 成立，则称 $\hat{\theta}_n$ 具有 Bahadur 渐近有效性. 简称 $\hat{\theta}_n$ 为 θ 的 B. A. E. 估计.

此处矩阵 $A_{n\times k}$ 的模 $\|A\|$ 定义为: 对于任何 k 维向量 y,

$$\|A\|=\max_{\|y\|=1}\|AY\|,$$

$\|Y\|$ 表示向量的欧氏模. 由线性代数知，如果 $A>0$, $\|A\|$ 为 A 的最大特征根.

又若 $g(\theta)$ 为具有一阶连续偏导数的 k 维 $(k\leqslant p)$ 向量函数，$\hat{g}_n$ 为其依赖于 $x_1, \cdots, x_n$ 的估计.

定义 4.2 若 $\hat{g}_n$ 为 $g(\theta)$ 的相合估计，如果对一切 $\theta\in\Theta$ 有下式成立:

$$\lim_{\varepsilon\to0}\sup\lim_{n\to\infty}\sup\frac{1}{n\varepsilon^2}\log P_\theta(\|\hat{g}_n-g(\theta)\|\geqslant\varepsilon)$$

$$\leqslant-\frac{1}{2\|(D(\theta))'I^{-1}(\theta)(D(\theta))\|},\qquad(4.2)$$

则称 $\hat{g}_n$ 为 $g(\theta)$ 的 B. A. E 估计. 此处 $D(\theta)=\left(\dfrac{\partial g(\theta)}{\partial\theta}\right)_{k\times p}$.

Bahadur 渐近有效性的定义并没有对估计类作多少限制，例如要求属于某阶相合估计，或要求具有某种渐近无偏性等. 因此这种有效性应用较广，当然比较的标准也就粗糙一些. 例如上节所举的超有效估计的例子(例 3.1), $\bar{x}$ 及

$$Q_n=\begin{cases}\dfrac{1}{2}\bar{x}, & \text{当 } |\bar{x}|\leqslant\log n/\sqrt{n};\\ \bar{x}, & \text{其它}.\end{cases}$$

在这里都是 θ 的 B. A. E 估计，即用 Bahadur 有效性的标准不认为它们之

间有差异.

对于定义 4.1 及 4.2, 人们自然要提出这样的问题: 为什么满足(4.1) 与(4.2)式的估计就称为"渐近有效"的? 它在大样本时的优良性的根据是什么? 下面我们来证明: 在一定条件下, (4.1)与(4.2)式的右端是 θ, $g(\theta)$的相合估计 $\hat{\theta}_n$, $\hat{g}_n$ 代入(4.1)及 (4.2) 式左边的值的下界. 这样就回答了上面提出的问题.

我们对分布族作如下假设:

1° $f(x, \theta)$, $\theta\in\Theta$ 有共同支撑, $\mu(f(x, \theta_1)\neq f(x, \theta_2))>0$, 当 $\theta_1\neq\theta_2$ $(\in\Theta)$.

2° $Z(x, \theta)=\dfrac{\partial \log f(x, \theta)}{\partial\theta}$ 对每一 x 存在有限, 且是 θ 的连续函数.

3° 对每个 $\theta_0\in\Theta$, 存在一个 θ_0 的邻域 $V_{\theta_0}\subset\Theta$ 及 $A(x, \theta_0)>0, B(x,\theta_0)>0$, 使得

$$\left|\frac{f(x, \theta)}{f(x, \theta_0)}-1\right|\leqslant A(x, \theta_0)|\theta-\theta_0| \quad (\text{当 } \theta\in V_{\theta_0});$$

$$|Z(x, \theta)|\leqslant B(x, \theta_0) \quad (\text{当 } \theta\in V_{\theta_0});$$

$$I(\theta_0)>0.$$

其中 $E_{\theta_0}A(x, \theta_0)<\infty$, $E_{\theta_0}B^2(x, \theta_0)<\infty$, $E_{\theta_0}A(x, \theta_0)B^2(x, \theta_0)<\infty$. 对待估函数假设:

4° $g(\theta)$是$k(\leqslant p)$维有连续一阶偏导数的向量函数.

定理4.1 如果分布族满足上述 1°~3° 的假设条件, $g(\theta)$ 满足 4°, 则对 θ 或 $g(\theta)$的任一相合估计 $\hat{\theta}_n$ 或 $\hat{g}_n$, 皆有

$$\liminf_{\varepsilon\to 0}\ \liminf_{n\to\infty}\frac{1}{n\varepsilon^2}\ln P_{\theta_0}(\|\hat{\theta}_n-\theta_0\|>\varepsilon)\geqslant-\frac{1}{2\|I^{-1}(\theta_0)\|}; \tag{4.3}$$

$$\liminf_{\varepsilon\to 0}\ \liminf_{n\to\infty}\frac{1}{n\varepsilon^2}\ln P_{\theta_0}(\|\hat{g}_n-g(\theta_0)\|>\varepsilon)$$
$$\geqslant-\frac{1}{2\|(D(\theta_0))'I^{-1}(\theta_0)(D(\theta_0))\|} \tag{4.4}$$

(对任一 $\theta_0\in\Theta$).

证 首先对 $g(\theta)$为连续可微实函数的情况来证明(4.4)式, 然后对 $g(\theta)$为一般 k 维向量函数来证(4.4)式. 至于(4.3)式, 它是(4.4)式的特殊情况. 不妨设 $D(\theta_0)\neq 0$, 因为若 $D(\theta_0)=0$ 时(4.4)式显然成立. 因此

$$(D(\theta_0))'I^{-1}(\theta_0)(D(\theta_0))>0.$$

记 $d=(D(\theta_0))'I^{-1}(\theta_0)(D(\theta_0))$.

选取 $0<\lambda<1$, 并取充分小的 $\varepsilon>0$, 使 $\theta^*=\theta_0+\varepsilon I^{-1}D\in\Theta$. 于是 $\varepsilon\to 0$ 时有

$$g(\theta^*)-g(\theta_0)=(D(\theta_0))'(\theta^*-\theta_0)+o(|\theta^*-\theta_0|)=\varepsilon d+o(\varepsilon). \tag{4.5}$$

若记 $\delta=\varepsilon d$，取 $\lambda\in(0,1)$，由此，当 ε 充分小时就有

$$g(\theta^*)-g(\theta_0)>\lambda\delta. \tag{4.6}$$

现在来考虑假设检验 $H_0:\theta=\theta_0$，$K:\theta=\theta^*$. 限定在 K 成立时接受假设 K 的概率为 $\frac{1}{2}$，由 Neyman-Pearson 基本引理，存在基本唯一的检验函数

$$\phi_n^*(x^{(n)})=\begin{cases}1, & \text{当 } r_n(x^{(n)})>C_n;\\ \delta_n, & \text{当 } r_n(x^{(n)})=C_n;\\ 0, & \text{当 } r_n(x^{(n)})<C_n.\end{cases} \tag{4.7}$$

其中，$x^{(n)}=(x_1,\cdots,x_n)$，

$$r_n(x^{(n)})=\prod_{i=1}^n f(x_i,\theta^*)\Big/\prod_{i=1}^n f(x_i,\theta_0),$$

C_n、δ_n 是与 $x^{(n)}$ 无关的常数，使得 $E_{\theta^*}\phi_n^*(X^{(n)})=\frac{1}{2}$，而 $\alpha_n\triangleq E_{\theta_0}\phi_n^*(X^{(n)})\leqslant E_{\theta_0}\phi_n(X^{(n)})$，这里 $\phi_n(X^{(n)})$ 为任一满足条件 $E_{\theta^*}(\phi_n(X^{(n)}))\geqslant\frac{1}{2}$ 的检验函数.

记 $$d_n=e^{nI(\theta^*,\theta_0)+n\varepsilon'}\quad(\varepsilon'>0).$$

$$\begin{aligned}\alpha_n=E_{\theta_0}\phi_n^*(X^{(n)})&\geqslant\int_{r_n<d_n}\phi_n^*\prod_{i=1}^n(f(x_i,\theta_0)d\mu(x_i))\\&\geqslant\frac{1}{d_n}\int_{r_n<d_n}\phi_n^*\prod_{i=1}^n(f(x_i,\theta^*)d\mu(x_i))\\&=\frac{1}{d_n}\left\{\frac{1}{2}-\int_{r_n>d_n}\phi_n^*\prod_{i=1}^n(f(x_i,\theta^*)d\mu(x_i))\right\}\\&\geqslant\frac{1}{d_n}\left\{\frac{1}{2}-P_{\theta^*}[r_n(X^{(n)})>d_n]\right\}.\end{aligned} \tag{4.8}$$

由大数法则，并注意 $E_{\theta^*}\log\frac{f(x,\theta^*)}{f(x,\theta_0)}=I(\theta^*,\theta_0)$，故

$$\begin{aligned}&\lim_{n\to\infty}P_{\theta^*}\{r_n(X^{(n)})>d_n\}\\&=\lim_{n\to\infty}P_{\theta^*}\{\log r_n(x^{(n)})>\log d_n\}\\&=\lim_{n\to\infty}P_{\theta^*}\left\{\frac{1}{n}\sum_{i=1}^n\log\frac{f(x_i,\theta^*)}{f(x_i,\theta_0)}>I(\theta^*,\theta_0)+\varepsilon'\right\}=0.\end{aligned} \tag{4.9}$$

由(4.8)及(4.9)就有

$$\liminf_{n\to\infty}\frac{1}{n}\log\alpha_n\geqslant-(I(\theta^*,\theta_0)+\varepsilon'). \tag{4.10}$$

由 $\varepsilon'>0$ 的任意性，推得

$$\liminf_{n\to\infty}\frac{1}{n}\log\alpha_n\geqslant-I(\theta^*,\theta_0).$$

又由本章§1引理1.1知：当 $\varepsilon\to 0$ 时，

$$\begin{aligned}I(\theta^*,\theta_0)&=\frac{1}{2}(\theta^*-\theta_0)'I(\theta_0)(\theta^*-\theta_0)(1+o(1))\\&=\frac{1}{2}\varepsilon^2(D(\theta_0))'I^{-1}(\theta_0)I(\theta_0)I^{-1}(\theta_0)(D(\theta_0))(1+o(1))\\&=\frac{1}{2}\varepsilon^2 d(1+o(1)).\end{aligned}\tag{4.11}$$

现在假设 $\tilde{g}_n$ 是 $g(\theta)$ 的任一相合估计．令

$$\phi_n(x^{(n)})=\begin{cases}1, & \text{当 } |\tilde{g}_n-g(\theta_0)|\geqslant\lambda\delta;\\0, & \text{其它.}\end{cases}\tag{4.12}$$

由于 $\tilde{g}_n$ 是相合估计，注意(4.6)式，有

$$\begin{aligned}\liminf_{n\to\infty}E_{\theta^*}\{\phi_n(X^{(n)})\}&=\liminf_{n\to\infty}P_{\theta^*}\{|\tilde{g}_n-g(\theta_0)|\geqslant\lambda\delta\}\\&\geqslant\liminf_{n\to\infty}P_{\theta^*}\{|g(\theta^*)-g(\theta_0)|-|\tilde{g}_n-g(\theta^*)|\geqslant\lambda\delta\}\\&=\liminf_{n\to\infty}P_{\theta^*}\{|\tilde{g}_n-g(\theta^*)|\leqslant|g(\theta^*)-g(\theta_0)|-\lambda\delta\}=1.\end{aligned}\tag{4.13}$$

故存在充分大的 N，使当 $n\geqslant N$ 时，

$$E_{\theta^*}\{\phi_n(X^{(n)})\}\geqslant\frac{1}{2}.$$

那么由 Neyman-Pearson 引理知

$$E_{\theta_0}\{\phi_n(X^{(n)})\}\geqslant\alpha_n.\tag{4.14}$$

于是由(4.10)，(4.11)，(4.14)，便有

$$\begin{aligned}&\liminf_{\varepsilon\to0}\liminf_{n\to\infty}\frac{1}{n\varepsilon^2}\log P_{\theta_0}\{|\tilde{g}_n-g(\theta_0)|\geqslant\varepsilon\}\\&=\liminf_{\varepsilon\to0}\liminf_{n\to\infty}\frac{1}{n(\lambda\delta)^2}\log P_{\theta_0}\{|\tilde{g}_n-g(\theta_0)|\geqslant\lambda\delta\}\\&\geqslant\liminf_{\varepsilon\to0}\liminf_{n\to\infty}\frac{1}{n(\lambda\delta)^2}\log\alpha_n\\&\geqslant\liminf_{\varepsilon\to0}\left\{-\frac{\left(\frac{1}{2}\right)\varepsilon^2 d}{(\lambda\delta)^2}\right\}=-\frac{1}{2\lambda^2 d}\quad(\because\quad\delta=\varepsilon d).\end{aligned}\tag{4.15}$$

再由 $\lambda\in(0,1)$ 的任意性，令 $\lambda\to1$，由上式得

$$\begin{aligned}&\liminf_{\varepsilon\to0}\liminf_{n\to\infty}\frac{1}{n\varepsilon^2}\log P_{\theta_0}\{|\tilde{g}_n-g(\theta_0)|\geqslant\varepsilon\}\\&\geqslant-\frac{1}{2\|(D(\theta_0))'I^{-1}(\theta_0)(D(\theta_0))\|}.\end{aligned}\tag{4.16}$$

这就是(4.4)式．为进一步证明当 $g(\theta)$ 为 k 维向量函数时上式成立，任取 $l\in R^k$，$\|l\|=1$，则 $l'g(\theta)$ 为一维实函数．设 $\tilde{g}_n$ 为 g 的任一相合估计，因为 $|l'(\tilde{g}_n-g(\theta_0))|\leqslant\|l\|\|\tilde{g}_n-g(\theta_0)\|=\|\tilde{g}_n-g(\theta_0)\|$，故容易看出 $l'\tilde{g}_n$ 是 $l'g(\theta)$

的相合估计. 由上面证明得

$$\liminf_{\varepsilon\to 0}\liminf_{n\to\infty}\frac{1}{n\varepsilon^2}P_{\theta_0}\{|l'(\tilde{g}_n-g(\theta_0))|\geqslant\varepsilon\}$$

$$=-\frac{1}{2l'(D(\theta_0))'I^{-1}(\theta_0)(D(\theta_0))l}. \tag{4.17}$$

对上式右端关于 l 取极大值, 再由 $\|A\|$ 的定义, 便推得 (4.4) 式. 定理证毕.

上面的定理首先由 Bahadur 给出, 我们对定理的条件有所减弱. 对于某些非共同支撑的情况, 参数估计的 Bahadur 渐近有效性在卢昆亮[18]文中已有讨论. 显然定义中的 Fisher 信息阵, 必然有相应的变化.

§5 Cramér 渐近有效性与矩估计的渐近方差

Cramér 于 1945 年在他的名著 [19] 中从估计与真值差的二阶矩达到极小出发, 提出了他的渐近有效估计的定义. 近三十年来不少作者认为有了渐近方差的概念就够了. 事实上, 这两个概念还是有差别的. 与 Cramér 渐近有效性有关的是建立在矩估计基础上的某些估计的渐近方差估计问题, 也是 Cramér 首先建立这方面的有关定理. 矩估计是经常采用的一种估计, 其方法虽较古老, 但有简单易行的优点, 在某些场合下, 具有很多优良性质.

仍假定 $X_1, X_2, \cdots$ 为 $iid.$ 随机序列, 它们具有分布族 $dP_\theta(x)=f(x,\theta)d\mu(x)$, $\theta\in\Theta$, Θ 是 R^p 中的开域. $\hat{\theta}_n(x_1,\cdots,x_n)$ 表示 θ 的依赖于 $x_1,\cdots,x_n$ 的估计. 分布族的 Fisher 信息矩阵存在且有限.

定义 5.1 若 θ 的相合估计 $\hat{\theta}_n$ 满足下列式子, 则称其为 θ 的 Cramér 渐近有效估计, 简记为 C. A. E 估计:

$$\lim_{n\to\infty} nE_\theta\{(\hat{\theta}_n-\theta)(\hat{\theta}_n-\theta)'\}=I^{-1}(\theta), \quad \theta\in\Theta. \tag{5.1}$$

定义 5.2 若 $\hat{g}_n$ 是 θ 的 $k(\leqslant p)$ 维向量函数 $g(\theta)$ 的相合估计. 如果对一切 $\theta\in\Theta$, 有下式成立:

$$\lim_{n\to\infty} nE_\theta\{(\hat{g}_n-g(\theta))(\hat{g}_n-g(\theta))'\}=(D(\theta))'I^{-1}(\theta)(D(\theta)). \tag{5.2}$$

则称 $\hat{g}_n$ 为 $g(\theta)$ 的 C. A. E. 估计.

为什么对满足上述条件的估计称之为具有渐近有效呢？这是因为在某类估计中，(5.1) 与 (5.2) 的右端为二次损失的渐近下界. 所谓渐近有效就是能达到此下界. 下面就来证明这一事实.

定理 5.1(Walker)　若分布族满足 § 4中所设的条件1°～3°，并进一步假设 θ 的估计 $\hat{\theta}_n$ 满足

i) $\left|\frac{\partial}{\partial\theta} b_n(\theta)\right| \triangleq \left|\frac{\partial}{\partial\theta} E_\theta(\hat{\theta}_n-\theta)\right| \to 0 \quad (n\to\infty)$,

则对一切 $\theta\in\Theta$ 有

$$\liminf_{n\to\infty} nE_\theta(\hat{\theta}_n-\theta)(\hat{\theta}_n-\theta)' \geqslant I^{-1}(\theta), \tag{5.3}$$

$I(\theta)$ 为 Fisher 信息阵.

如果对一切 $\theta\in\Theta$，进一步假设

ii) $\mathscr{L}_\theta(\sqrt{n}(\hat{\theta}_n-\theta))\to N(0, A(\theta))$；

iii) $n|\hat{\theta}_n-\theta|^2$ 关于 $\{P_{n,\theta}(x_1, \cdots, x_n)\}$ 一致可积，则对每一个 $\theta\in\Theta$，有

$$A(\theta)\geqslant I^{-1}(\theta).$$

证　根据对分布族的假设，不难验证 C-R 不等式条件成立. 故对任意 $\xi\in R^p$，

$$n\xi' E_\theta\{(\hat{\theta}_n-\theta)(\hat{\theta}_n-\theta)'\}\xi$$

$$\geqslant \xi' I^{-1}(\theta)\xi\left[1+\frac{\xi'\left(\frac{\partial}{\partial\theta}b_n(\theta)\right)I^{-1}(\theta)\xi}{\xi' I^{-1}(\theta)\xi}\right]^2. \tag{5.4}$$

故由假设 i) 有

$$\liminf_{n\to\infty} n\xi' E_\theta\{(\hat{\theta}_n-\theta)(\hat{\theta}_n-\theta)'\}\xi \geqslant \xi' I^{-1}(\theta)\xi. \tag{5.5}$$

由 ξ 的任意性，由(5.5)式立即可推得 (5.3) 式. 从而定理的第一部分得证. 至于第二部分，由所给条件 ii) 及 iii)，利用 Helly-Bray 定理知，(5.5) 式左端即为 $A(\theta)$，故有 $A(\theta)\geqslant I^{-1}(\theta)$. 定理证毕.

系 5.1　如果上述定理中的条件 iii) 换为对于 θ 的具有一阶

连续偏导数的 $k(\leqslant p)$ 维向量函数 $g(\theta)$ 的估计 $\hat{g}_n$ 满足条件：$n|\hat{g}_n-g(\theta)|^2$ 一致可积，则有

$$(D(\theta))'A(\theta)(D(\theta))\geqslant(D(\theta))'I^{-1}(\theta)(D(\theta)) \qquad (5.6)$$

对一切 $\theta\in\Theta$ 成立.

此系的证明很显然，从略.

注 1°) 定理中的条件 i) 是对估计有偏性的一种限制. 如果 $\hat{\theta}_n$ 就是 θ 的无偏估计，则条件自然成立. 如果此条件不成立，则定理的结论难以保证成立. §3 中开始举的超有效的例子，就是不满足此条件而产生了超有效的情况.

2°) 此定理也告诉我们，如果 $\theta, g(\theta)$ 的估计 $\hat{\theta}_n, \hat{g}_n$ 是 B.A.N 估计，只要 $n|\hat{\theta}_n-\theta|^2$ 或 $n|\hat{g}_n-g(\theta)|^2$ 一致可积，则它也是 C.A.E. 估计. 由于 Helly-Bray 定理的必要性也成立（见 Loéve, «Probability Theory»），因此，一致可积性是一个 B.A.N 估计为 C.A.E. 估计的充要条件. 不过一致可积性一般难以验证，这是证明 C.A.E. 估计的困难所在.

下面来讨论矩估计的渐近性质. 有不少情况是：要估计的未知参数或其函数是描述总体的某些特征，而这些特征往往是总体的各阶矩的函数. 例如均值、协方差阵、相关阵、变异系数等等. 在我们估计这些未知参数或其函数时，就采用相应的样本的各阶矩的函数作为估计. 设 $X_1, X_2, \cdots$ 为 *iid.* 随机序列. 那么

$$\overline{X}_n=\frac{1}{n}\sum_{i=1}^{n}X_i;$$

$$\hat{A}_n=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X}_n)(X_i-\overline{X}_n)';$$

$$\hat{R}_n=(\hat{r}_{k\nu}^{(n)})$$

分别为母体的均值，协方差阵，相关阵的矩估计. 这里

$$\hat{r}_{k\nu}^{(n)}=\sum_{i=1}^{n}(x_{ik}-\bar{x}_{k,n})(x_{i\nu}-\bar{x}_{\nu,n})\Big/\sqrt{\sum_{i=1}^{n}(x_{ik}-\bar{x}_{k,n})^2\sum_{i=1}^{n}(x_{i\nu}-\bar{x}_{\nu,n})^2},$$

而 x_{ik} 是 x_i 的第 k 个分量；$\bar{x}_{n,k}$ 为 $\bar{x}_n$ 的第 k 个分量. 在我们要估

计的是这些矩的函数时，只要把样本相应的矩代入相应的待估函数，便得到矩估计. 比如，要估计变异系数 σ/μ (σ 为母体的标准差，μ 为均值)，其矩估计为 $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x}_n)^2}\Big/\bar{x}_n$; 再如打靶要估计击中某目标(如圆、方块等)的命中率

$$\Phi(B;\mu,A)=\frac{1}{2\pi|A|^{\frac{1}{2}}}\int_B e^{-\frac{1}{2}(t-\mu)'A^{-1}(t-\mu)}dt,$$

这里假定弹着点服从二维正态分布 $N(\mu, A)$，B 是目标区域，此时矩估计为 $\Phi(B;\overline{X}_n,\hat{A}_n)$.

在此应指出：如果 X_1 服从指数分布族，其分布密度为 $dP_\theta(x)=\beta(\theta)e^{\theta' x}d\mu(x)$，$\theta\in\Theta$. 当样本的大小为 n 时，$\overline{X}_n$ 是 θ 的完全充分统计量. 由充分统计量的充分性原则(见第一章§6)知，待估的 θ 的向量函数 $\alpha(\theta)$ 的优良估计 $\hat{\alpha}_n$ 是 $\overline{X}_n$ 的函数. 这是一个矩估计. 这里自然产生一个如何估算 $\hat{\alpha}_n$ 的均值及二次损失 $E(\hat{\alpha}_n-\alpha)(\hat{\alpha}_n-\alpha)'$ 的问题. 估计矩的函数也同样存在这样的问题. Cramér 在其名著 [19] 中研究了这个问题. 下面我们采用 [21]，介绍更一般形式的结果.

定理 5.2 设 $X_1, X_2, \cdots$ 为 *iid.* 随机序列. r 维向量 $\alpha=(\alpha_1,\cdots,\alpha_r)'$ 是 $q(\leqslant r)$ 维向量 $C=(c_1,\cdots,c_q)'$ 的函数. 若 $\hat{C}_n$ 是 C 的估计量. 假设

i) $\alpha(C)$ 在 C 的某个邻域内有二阶连续偏导数;

ii) 存在常数 $\delta_1>0$, $\delta_2\geqslant 0$, $M>0$, 使得

$$E|\alpha(\hat{C}_n)|^{2+\delta_1}\leqslant Mn^{\delta_2};$$

iii) 存在正整数 $K_0\geqslant(3+1.5\delta_1+2\delta_2)\delta_1^{-1}$, 使得

$$E(\hat{C}_n-C)=O(n^{-1}),$$

$$E|\hat{C}_n-C|^{2k}=O(n^{-k})\quad (\text{当 } K<K_0+1).$$

那么 $\hat{\alpha}_n=\alpha(\hat{C}_n)$ 作为 α 的估计时，差 $\tilde{\alpha}_n=\hat{\alpha}_n-\alpha$ 有下列性质:

1°) $E\tilde{\alpha}_n=O(n^{-1})$; (5.7)

2°) $E\tilde{\alpha}_n\tilde{\alpha}_n'=\left(\frac{\partial\alpha}{\partial C}\right)E(\hat{C}_n\hat{C}_n')\left(\frac{\partial\alpha}{\partial C}\right)'+O(n^{-\frac{3}{2}})$, (5.8)

这里 $\widetilde{C}_n=\hat{C}_n-C$;

3°) 若 $\mathscr{L}(\sqrt{n}(\hat{C}_n-C))\to N(0, A)$, 则

$$\mathscr{L}(\sqrt{n}\,\tilde{\alpha}_n)\to N\left(0,\left(\frac{\partial\alpha}{\partial C}\right)A\left(\frac{\partial\alpha}{\partial C}\right)'\right). \tag{5.9}$$

证 根据假设 iii), $\hat{C}_n$ 是 C 的相合估计. 又由 Taylor 展开式

$$\tilde{\alpha}_n=\hat{\alpha}_n-\alpha=\left(\frac{\partial\alpha}{\partial C}\right)(\hat{C}_n-C)+o(|\hat{C}_n-C|).$$

故 $\mathscr{L}(\sqrt{n}\,\tilde{\alpha}_n)$ 的极限分布与 $\mathscr{L}\left(\sqrt{n}\left(\frac{\partial\alpha}{\partial C}\right)(\hat{C}_n-C)\right)$ 的极限分布相同. 由于 $\mathscr{L}(\sqrt{n}\,(\hat{C}_n-C))\to N(0, A)$ (当 $n\to\infty$), 故 $\mathscr{L}\left(\sqrt{n}\left(\frac{\partial\alpha}{\partial C}\right)(\hat{C}_n-C)\right)\to N\left(0,\left(\frac{\partial\alpha}{\partial C}\right)A\left(\frac{\partial\alpha}{\partial C}\right)'\right)$, 从而定理的结论 3°)成立.

设 $\varepsilon>0$. 记 $D_n=\{(x_1, \cdots, x_n)': |\widetilde{C}_n|<\varepsilon\}$, D_n^c 为 D_n 的余集. 取 $\max(2, k_0)\leqslant l<k_0+1$, $l\geqslant(2+\delta_1+\delta_2)/(1+\delta_1)$. 由 Chebyshev 不等式及假设 iii)知, 存在常数 $M_l>0$, 使

$$P(D_n^c)\leqslant\frac{E|\widetilde{C}_n|^{2l}}{\varepsilon^{2l}}\leqslant M_l\varepsilon^{-2l}n^{-l}. \tag{5.10}$$

令 $P_n(t)$ 表示 $\hat{C}_n$ 的联合分布, 于是

$$E\alpha(\hat{C}_n)=\int_{D_n}\alpha(t)dP_n(t)+\int_{D_n^c}\alpha(t)dP_n(t). \tag{5.11}$$

根据 Hölder 不等式

$$\begin{aligned}\left|\int_{D_n^c}\alpha(t)dP_n(t)\right| &\leqslant\left\{\int_{D_n^c}|\alpha(t)|^{2+\delta_1}dP_n\right\}^{\frac{1}{2+\delta_1}}\left\{\int_{D_n^c}dP_n\right\}^{\frac{1+\delta_1}{2+\delta_1}}\\ &\leqslant\{E|\alpha(\hat{C}_n)|^{2+\delta_1}\}^{\frac{1}{2+\delta_1}}\{P(D_n^c)\}^{\frac{1+\delta_1}{2+\delta_1}}\\ &\leqslant M^{\frac{1}{2+\delta_1}}M_l^{\frac{1+\delta_1}{2+\delta_1}}\varepsilon^{-2l\frac{1+\delta_1}{2+\delta_1}}n^{\frac{\delta_2-l(1+\delta_1)}{2+\delta_1}}=O(n^{-1}).\end{aligned} \tag{5.12}$$

取 ε 充分小, 使 $\alpha(C)$ 在 C 的 ε-邻域内存在二阶连续偏导数. 依 Taylor 展开式

$$\alpha(\hat{C}_n)=\alpha(C)+\left(\frac{\partial\alpha}{\partial C}\right)\widetilde{C}+R, \tag{5.13}$$

其中 $R=(R_1, \cdots, R_r)'$,

$$R_i=\frac{1}{2}\sum_{k,l}^{q}\frac{\partial^2\alpha_i}{\partial C_k\partial C_l}\bigg|_{C=C^*}\widetilde{C}_k\widetilde{C}_l \quad (i=1, \cdots, r),$$

而 C^* 为 C 与 $\hat{C}_n$ 之间的某个点.

由条件 i)知,存在 s(只与 C 有关)使

$$\left|\sum_{k,l=1}^{q}\frac{\partial^2\alpha_i}{\partial C_k\partial C_l}\right|\leqslant s \quad (\text{当 } |\widetilde{C}_n|<\varepsilon). \tag{5.14}$$

而

$$\int_{D_n}\alpha(t)dP_n(t)=\alpha(C)P_n(D_n)+\frac{\partial\alpha}{\partial C}\int_{D_n}\widetilde{C}_n dP_n+\int_{D_n}R\,dP_n; \tag{5.15}$$

$$\int_{D_n}\widetilde{C}_n dP_n=E\widetilde{C}_n-\int_{D_n^c}\widetilde{C}_n dP_n=O(n^{-1})-\int_{D_n^c}\widetilde{C}_n dP_n. \tag{5.16}$$

由 Hölder 不等式,

$$\left|\int_{D_n^c}\widetilde{C}_n dP_n\right|\leqslant\{E|\widetilde{C}_n|^2\}^{\frac{1}{2}}\{P_n(D_n^c)\}^{\frac{1}{2}}$$

$$=O(n^{-\frac{1}{2}})M_l^{\frac{1}{2}}\varepsilon^{-l}n^{-\frac{l}{2}}=O(n^{-1}); \tag{5.17}$$

$$\left|\int_{D_n}R_i dP_n\right|\leqslant\frac{1}{2}S\int|\widetilde{C}_n|^2 dP_n=O(n^{-1}) \quad (i=1, \cdots, r). \tag{5.18}$$

由(5.13)及(5.15)～(5.18), 就有

$$\int_{D_n}\alpha(t)dP_n(t)=\alpha(C)P_n(D_n)+O(n^{-1})$$

$$=\alpha-\alpha P_n(D_n^c)+O(n^{-1})=\alpha+O(n^{-1}). \tag{5.19}$$

再利用(5.11)、(5.12)、(5.19), 就有

$$E\alpha(\hat{C}_n)=\alpha+O\left(\frac{1}{n}\right).$$

这证明了定理的结论 1°).

现在证明 2°): 由(5.13),

$$
\begin{aligned}
E\tilde{\alpha}(\hat{C}_n)\tilde{\alpha}(\hat{C}_n)' = & \int_{D_n}\tilde{\alpha}(t)\tilde{\alpha}'(t)\,dP_n(t) + \int_{D_n^c}(\tilde{\alpha}(t))(\tilde{\alpha}(t))'\,dP_n(t) \\
& - \int_{D_n}\left(\frac{\partial\alpha}{\partial C}\right)\tilde{C}_n\tilde{C}_n'\left(\frac{\partial\alpha}{\partial C}\right)'dP_n \\
& + \int_{D_n}RR'\,dP_n + \int_{D_n}\frac{\partial\alpha}{\partial C}\tilde{C}_nR'\,dP_n \\
& + \int_{D_n}R\tilde{C}_n'\left(\frac{\partial\alpha}{\partial C}\right)'dP_n + \int_{D_n^c}\tilde{\alpha}\tilde{\alpha}'\,dP_n. \qquad (5.20)
\end{aligned}
$$

由(5.10)及假设条件 ii)、iii)，

$$
\begin{aligned}
& \left|\int_{D_n^c}(\tilde{\alpha}(t))(\tilde{\alpha}(t))'\,dP_n\right| \leqslant \int_{D_n^c}|\tilde{\alpha}\tilde{\alpha}'|\,dP_n \\
& \leqslant \int_{D_n^c}|\tilde{\alpha}(t)|^2\,dP_n \leqslant \left\{\int|\tilde{\alpha}|^{2+\delta_1}\,dP_n\right\}^{\frac{2}{2+\delta_1}}\left\{\int_{D_n^c}dP_n\right\}^{\frac{\delta_1}{2+\delta_1}} \\
& \leqslant Mn^{\frac{2\delta_1}{2+\delta_1}}M_l s^{-\frac{2\delta_1K_n}{2+\delta_1}}n^{-\frac{K_n\delta_1}{2+\delta_1}} = O\left(n^{-\frac{3}{2}}\right). \qquad (5.21)
\end{aligned}
$$

同样由(5.10)及假设条件 ii)、iii)，

$$
\begin{aligned}
\int_{D_n}\left(\frac{\partial\alpha}{\partial C}\right)(\tilde{C}_n\tilde{C}_n')\left(\frac{\partial\alpha}{\partial C}\right)'dP_n = & \frac{\partial\alpha}{\partial C}E(\tilde{C}_n\tilde{C}_n')\left(\frac{\partial\alpha}{\partial C}\right)' \\
& - \int_{D_n^c}\left(\frac{\partial\alpha}{\partial C}\right)(\tilde{C}_n\tilde{C}_n')\left(\frac{\partial\alpha}{\partial C}\right)'dP_n. \qquad (5.22)
\end{aligned}
$$

而

$$
\begin{aligned}
& \left|\int_{D_n^c}\left(\frac{\partial\alpha}{\partial C}\right)(\tilde{C}_n\tilde{C}_n')\left(\frac{\partial\alpha}{\partial C}\right)'dP_n\right| \\
& \leqslant q^2r^2\int_{D_n^c}|\tilde{C}_n|^2\,dP_n\cdot\left|\frac{\partial\alpha}{\partial C}\right|^2 \\
& \leqslant q^2r^2(E\tilde{C}_n^4)^{\frac{1}{2}}\{P(D_n^c)\}^{\frac{1}{2}}\left|\frac{\partial\alpha}{\partial C}\right|^2 = O\left(n^{-\frac{1+K_n}{2}}\right). \qquad (5.23)
\end{aligned}
$$

由(5.14)推得

$$
|R| \leqslant S|\tilde{C}_n|^2. \qquad (5.24)
$$

因此由(5.24)、Hölder 不等式、假设条件 iii)，有

$$\left|\int_{D_n} R\widetilde{C}_n'\left(\frac{\partial\alpha}{\partial C}\right)' dP_n\right| \leqslant rP\int |R|\,|\widetilde{C}_n|\,\left|\frac{\partial\alpha}{\partial C}\right| dP_n$$

$$\leqslant \mathrm{rps}\left|\frac{\partial\alpha}{\partial C}\right|\int |\widetilde{C}_n|^3\, dP_n.$$

$$\leqslant \mathrm{rps}\left|\frac{\partial\alpha}{\partial C}\right|\left\{\int |\widetilde{C}_n|^4\, dP_n\right\}^{\frac{3}{4}} = O(n^{-\frac{3}{2}}). \tag{5.25}$$

同理

$$\left|\int_{D_n} \frac{\partial\alpha}{\partial C}\widetilde{C}_n R'\, dP_n\right| = O(n^{-\frac{3}{2}}). \tag{5.26}$$

再利用(5.24)，就有

$$\left|\int_{D_n} RR'\, dP_n\right| \leqslant rs\int |\widetilde{C}_n|^4 dP_n = O(n^{-2}). \tag{5.27}$$

综合(5.20)～(5.23)，(5.25)～(5.27)，就有

$$E(\widetilde{\alpha}(\hat{C}_n))(\widetilde{\alpha}(\hat{C}_n))' = \left(\frac{\partial\alpha}{\partial C}\right)E(\widetilde{C}_n\widetilde{C}_n')\left(\frac{\partial\alpha}{\partial C}\right)' + O(n^{-\frac{3}{2}}). \tag{5.28}$$

因此证得结论 3°. 至此，定理完全证毕.

注 如果定理条件 iii) 进一步假设 $E(\widetilde{C}_{nk}\widetilde{C}_{nl}\widetilde{C}_{nm}) = O(n^{-2})$ $(k, l, m=1, \cdots, q)$，$\widetilde{C}_{nk}$ 为 $\widetilde{C}_n$ 的第 k 个分量 $(k=1, \cdots, q)$. 适当放大 k_0，并把 i) 加强为三阶连续偏微分存在. 那么用同样的办法，把 Taytor 展开式展成三项，于是结论 2°) 可加强为 $O(n^{-2})$.

引理 5.1 设 $X_1, X_2, \cdots$ 为 iid. 随机序列，且 $E|X_1|^{k_0}<\infty$，则当 $k\leqslant k_0$ 时，有

$$E(\overline{X}_n - EX_1)^k = \begin{cases} (k-1)!!\,(D(X_1))^{\frac{k}{2}} n^{-\frac{k}{2}} + O(n^{-\frac{k}{2}-1}) & (\text{当 } k \text{ 为偶数}); \\ \dfrac{k-1}{6}\,k!!\,(D(X_1))^{\frac{k-3}{2}} E(X_1 - EX_1)^3 n^{-\frac{k+1}{2}} + O(n^{-[\frac{k}{2}]-2}) & (\text{当 } k \text{ 为奇数}). \end{cases} \tag{5.29}$$

此处 $\overline{X}_n = \dfrac{1}{n}\sum\limits_{i=1}^{n} X_i$.

证 首先由独立性，得

$$E[(X_1-EX_1)^{r_1}\cdots(X_n-EX_1)^{r_n}]=\prod_{i=1}^{n}E(X_i-EX_1)^{r_i}=0.$$

(5.30)

如果 $\sum_{i=1}^{n}r_i=k$, 且 $r_1, \cdots, r_n$ 中有一个为1. 利用上式, 并注意 $X_1, \cdots, X_n$ 为同分布的, 若记 $Y_i=X_i-EX_1$, 有

$$\begin{aligned}
E(\overline{X}_n-EX_1)^k&=\frac{1}{n^k}E\left(\sum_{i=1}^{n}Y_i\right)^k\\
&=\frac{k!}{n^k}\sum_{\substack{r_1+\cdots+r_n=k\\ r_\nu\geqslant 0,\nu=1,\cdots,n}}\frac{1}{r_1!\cdots r_n!}\prod_{i=1}^{n}E(Y_i^{r_i})\\
&=\frac{k!}{n^k}\sum_{j=1}^{\left[\frac{k}{2}\right]}\sum_{1\leqslant i_1<\cdots<i_j\leqslant n}\sum_{\substack{r_{i_1}+\cdots+r_{i_j}=k\\ r_{i_\nu}\geqslant 2,\nu=1,\cdots,j}}\frac{1}{r_{i_1}!\cdots r_{i_j}!}\prod_{\nu=1}^{j}EY_{i_\nu}^{r_{i_\nu}}\\
&=\frac{k!}{n^k}\binom{n}{\left[\frac{k}{2}\right]}2^{-\left[\frac{k}{2}\right]}3^{-k+2\left[\frac{k}{2}\right]}\left(\left[\frac{k}{2}\right]\right)^{+k-2\left[\frac{k}{2}\right]}\\
&\quad\times(EY_1^2)^{3\left[\frac{k}{2}\right]-k}(EY_1^3)^{k-2\left[\frac{k}{2}\right]}\\
&\quad+\frac{k!}{n^k}\sum_{j=1}^{\left[\frac{k}{2}\right]-1}\sum_{1\leqslant i_1<\cdots<i_j\leqslant n}\sum_{\substack{r_{i_1}+\cdots+r_{i_j}=k\\ r_{i_\nu}\geqslant 2,\nu=1,\cdots,j}}\\
&\quad\times\frac{1}{r_{i_1}!\cdots r_{i_j}!}\prod_{\nu=1}^{j}EY_{i_\nu}^{r_{i_\nu}}\\
&\triangleq I(n,k)+R(n,k);
\end{aligned}\tag{5.31}$$

$$\begin{aligned}
I(n,k)&=\begin{cases}\dfrac{k!}{n^k}\dbinom{n}{\left[\frac{k}{2}\right]}(D(X_1))^{\frac{k}{2}}2^{-\frac{k}{2}} & (\text{当 } k \text{ 为偶数})\\[2ex]
\dfrac{k!}{n^k}\dbinom{n}{\left[\frac{k}{2}\right]}\left[\dfrac{k}{2}\right](D(X_1))^{\left[\frac{k}{2}\right]-1}E(X_1-EX_1)^3 2^{-\left[\frac{k}{2}\right]}3^{-1} & (\text{当 } k \text{ 为奇数})\end{cases}\\
&=\begin{cases}\dfrac{(k-1)!!}{n^{k/2}}(D(X_1))^{\frac{k}{2}}+O\left(n^{-\frac{k}{2}-1}\right) & (\text{当 } k \text{ 为偶数})\\[2ex]
\dfrac{k!!}{6n^{\frac{k+1}{2}}}(k-1)(D(X_1))^{\frac{k-3}{2}}E(X_1-EX_1)^3+O\left(n^{-k+\left[\frac{k}{2}\right]-1}\right) & (\text{当 } k \text{ 为奇数}).\end{cases}
\end{aligned}\tag{5.32}$$

$$
|R(n,\ k)| \leqslant \frac{k!}{n^k} \sum_{j=1}^{\left[\frac{k}{2}\right]-1} \binom{n}{j} \sum_{\substack{r_1+\cdots+r_j=k \\ r_\nu \geqslant 2,\ \nu=1,\cdots,j}} \frac{1}{r_1!\cdots r_j!} \prod_{\nu=1}^{j} E|X_1|^{r_\nu}
$$

$$
= O\left(n^{-k-1+\left[\frac{k}{2}\right]}\right). \tag{5.33}
$$

由(5.31)～(5.33)，引理得证.

下面我们来说明，在指数分布族中用矩估计法所得到的对参数函数的估计是 C. A. E. 估计. 这是因为对指数族

$$
dP_\theta(x) = f(x,\ \theta)d\mu = \beta(\theta)e^{\theta' x} \quad (\theta \in \Theta)
$$

若 Θ 是开集，那么由第一章中的定理 2.2 知，X_1 的各阶矩都存在有限，而 $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ 是 $m(\theta) = E_\theta X_1$ 的 UMVUE，它的协方差达到 C-R 下界 $\frac{1}{n} I^{-1}(\theta)$. 因此，$\overline{X}_n$ 是 Cramér 意义下的渐近有效估计，也是 B. A. N. 估计. 如果我们估计的 $m(\theta)$ 的向量函数 $G(m)$（注意，下一章将要证明 θ 与 $m(\theta)$ 是 1-1 对应的，因此 θ 的向量函数可表为 $m(\theta)$ 的向量函数)，只要存在 $\delta > 0$，使得

$$
E_\theta |G(\overline{X}_n)|^{2+\delta} \leqslant M n^{\delta_2} \quad (\delta_2 \geqslant 0),
$$

那么由定理 5.2，就有

$$
E_\theta G(\overline{X}_n) = G(m) + O(n^{-1}). \tag{5.34}
$$

$$
E_\theta (G(\overline{X}_n) - G(m))(G(\overline{X}_n) - G(m))'
$$

$$
= \left(\frac{\partial G}{\partial m}\right) \mathrm{Var}\,(\overline{X}_n) \left(\frac{\partial G}{\partial m}\right)' + O\left(n^{-\frac{3}{2}}\right). \tag{5.35}
$$

因此 $G(m)$ 的矩估计 $G(\overline{X}_n)$ 也是 C. A. E. 估计. 在下章我们要证明在多数情况下，$\overline{X}_n$ 是 m 的极大似然估计，$G(\overline{X}_n)$ 也是 $G(m)$ 的极大似然估计. 由于指数分布族满足前面"注"中所提出的加强条件，因此(5.35)右端里的 $O(n^{-\frac{3}{2}})$ 可改为 $O(n^{-2})$.

问题与习题

1. 令随机变量 X 服从二项分布 $B(p,\ n)(0<p<1)$. 试求 $P\left(\left|\frac{X}{n} - p\right| > \varepsilon\right)$ 的 Chernoff 不等式.

2.* 在定理2.3中,如果 $\varepsilon_i \sim N(0, \sigma^2)$, 则定理的结论改为强相合也成立.

3. 令 $x_1, x_2, \cdots$ 为 $iid.$ 随机序列, 具有密度函数 $\frac{1}{\sigma} f\left(\frac{x-\theta}{\sigma}\right)$, 此处 $\theta \in R^1$ 与 $\sigma > 0$ 皆未知, 且当 $x \in R^1$ 时, $f(x) > 0$. $f'(x)$处处连续, 并且

$$I = \int_{-\infty}^{\infty} \frac{(f'(x))^2}{f(x)} dx < \infty.$$

试证明:

i) $x^*_{[pn]} - C_p = O_p(n^{-\frac{1}{2}})$, 此处 C_p 为分布函数 $F(x) = \int_{-\infty}^{x} f(t) dt$ 的 p 分位点. $x^*_{[pn]}$ 为 $x_1, x_2, \cdots, x_n$ 中第$[pn]$个次序统计量;

ii) 通过 $x^*_{[\frac{1}{4}n]}$、$x^*_{[\frac{n}{2}]}$、$x^*_{[\frac{3}{4}n]}$, 如何构造一个 θ 的渐近有效估计?

4. 在什么条件下? 回归模型中回归系数的最小二乘估计可改造为渐近有效估计.

5.* 在题3中, 若 $\sigma = 1$, $f(x) = I_{(-\frac{1}{2}, \frac{1}{2})}(x)$, 试求 θ 的渐近有效估计.

6. 设 $E|X|^\alpha < \infty (\alpha > 0)$, 则存在实函数 $g(t)$, 且 $\lim\limits_{t\to\infty} g(t) = \infty$, 使得 $E(|X|^\alpha g(|X|)) < \infty$.

[提示: 取 $g(t) = -\int_t^{\infty} x^\alpha dG(x)$, $G(x) = P(|X| > x)$.]

7.* 令 $x_1, x_2, \cdots, x_n$ 为 n 个 $iid.$ 随机变量, $EX_1 = 0$, $EX_1^{2k} < \infty$, 且 x_1 的分布关于原点对称. 则

$$E\left(\sum_{i=1}^{n} a_i x_i\right)^{2k} \leqslant 3(2k-1)!!\left(\sum_1^n a_i^2\right)^k Ex_1^{2k}.$$

[提示: 用多项式展开, 消去期望值为0的项, 放大 $Ex_{i_1}^{r_{i_1}} \cdots Ex_{i_j}^{r_{i_j}}$ 为 Ex_1^{2k} 以及相应的系数, 使之能写为$\left(\sum\limits_{i=1}^{n} a_i^2\right)^k$ 的形式.]

8. 令 $x_1, x_2, \cdots$ 为 $iid.$ 随机序列, 具有正态分布 $N(\theta, \sigma^2)$, $\theta \in R^1$, $\sigma^2 > 0$ 皆未知, 记

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \hat{\sigma}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{X}_n)^2.$$

试证明 $E\left[\Phi\left(-\frac{\overline{X}_n}{\hat{\sigma}_n}\right) - \Phi\left(-\frac{\theta}{\sigma}\right)\right]^2 = \frac{1}{n}\phi^2\left(\frac{\theta}{\sigma}\right)\left(\frac{\theta^2}{2\sigma^2} + 1\right) + O(n^{-2})$,

而且 $\Phi\left(-\frac{\overline{X}_n}{\hat{\sigma}_n}\right)$是 $\Phi\left(-\frac{\theta}{\sigma}\right)$的 Cramér 渐近有效估计, 此处 $\Phi(x)$ 与 $\phi(x)$ 分别表示标准正态分布和密度.

9. 令 $x_1, x_2, \cdots$为 $iid.$ 随机变量序列, 它具有密度函数 $\frac{1}{b-a}\left[\Phi\left(\frac{x-a}{\sigma}\right)\right.$

$-\Phi\left(\frac{x-b}{\sigma}\right)\Big]$，此处 $b>a$, $\sigma>0$ 皆未知. $\Phi(x)$ 同上，试求 x_1 的一到四阶的中心矩，从而构造出 a, b, σ^2 的矩估计，并求其渐近协方差阵.

10. 在定理3.3中，试求 $\lim\limits_{\varepsilon\to 0}\lim\limits_{n\to\infty}\frac{1}{n\varepsilon}\log P\{|\hat{\theta}_n-\theta|>\varepsilon\}$ 的值.

参 考 文 献

[1] M. Loeve: Probability Theory, D. Van. Nostrand, New York (1963).

[2] H. Chernoff: A measure of asymptotic efficiency for test of a hypotheses based on the sum of observations, Ann. Math. Statist. 23: 493～547 (1952).

[3] E. L. Lehmann: Test Statistical Hypotheses, John. Wiley, New York, (1959).

[4] F. Eicker: Limit theorem for regressions with unequal and dependent errors. Proc. Fifth. Berkeley Symp. Math. Statist. Prob. 1. pp. 59.

[5] H. Drygas: Weak and strong consistency of the least squares estimators in regression models, Z. Wahrscheinlichkeistheorie und Verw. Gebiete 34: 119～127 (1976).

[6] 陈希孺: i) Consistency of least squares estimates of regression coefficients in linear models, Kexue Tongbao, 6: 241～243 (1979).

ii) Consistency of least squares estimates in linear models Scientia Sinica, Special Issue 2: 162-176 (1979).

iii) «线性模型中误差方差估计的相合性», 中国科学, 11:1039～1050 (1979).

[7] T. W. Anderson and J. B. Taylor: Strong consistency of least squares estimates in normal linear regression, Ann. Statist. 4: 788～790 (1976).

[8] T. L. Lai, H. Robbins and C. Z. Wei: Strong consistency of least squares estimates in multiple regression II, J. Multivariate Anal. 9: 343～362 (1979).

[9] 陈桂景, T. L. Lai and C. Z. Wei: Convergence systems and strong consistency of least squares estimates in regression models. J. Multivariate Anal. 11:319～333 (1981).

[10] 赵林成: «线性模型中误差方差估计强相合的充要条件» 中国科学, 第10期, 1187～1191 (1981).

[11] R. Beran: A efficient and robust adaptive estimator of location, Ann. Statist. 6: 292～313 (1978).

[12] C. P. Sidney, and C. J. Stone: Fisher information and the Pitman estimator of a location parameter, Ann. Statist. 2: 225～247 (1974).

[13] E. Parzen: On estimation of a probability density function and mode, Ann. Math. Statist. 33: 1065～1076 (1962).

[14] 陈希孺: MSE propeperties of nearest neighbor density estimates, 中国科学,第 12 期, 1419～1428 (1981).

[15] R. R. Fisher: On the mathematical fundation of theoretical statistics, Philes Trans. Roy. Soc. A: 222: 309～324 (1922).

[16] 竹内启:《统计的推定の渐近理论》,教育出版株式会社 (1974).

[17] R. R. Bahadur: Rate of convergence of estimates and test statistics, Ann. Math. Statist. 38: 303～324, (1967).

[18] 卢昆亮:《截断参数估计的收敛速度》,科学通报,第 1 期, 11～13 (1982).

[19] H. Cramér: Mathematical Methods In Statistic, Princeton (1948).

[20] A. M. Walker: A note on asymptotic efficiency of an asymptotically normal estimator sequence, J. Roy. Stat. Soc. B 25 193～200 (1963).

[21] 安鸿志、成平:《自回归参数估计的渐近均方差》,应用数学学报, vol 3, No: 1, 13～33 (1980).

[22] J. Hajek: Local asymptotic minimax and admissibility in estimation, Proc. Sixth Berkely Symp. Math. Statist. Prob. 1 531～546 (1961).

[23] S. Zacks: The Theory Of Statistical Inference, John. Wiley & Sons. Inc. New York (1971).

第六章　极大似然估计

极大似然估计方法，是一种传统的、比较直观的参数估计方法，它有不少优良性质，至今仍然为数理统计研究者所感兴趣的主题之一，特别是关于它的渐近性质的研究. 本章介绍极大似然估计的定义和不变性、指数族的极大似然估计、极大似然估计的相合性、渐近有效性以及把非渐近有效估计改造为渐近有效的方法等.

§1　极大似然估计的定义及不变性

设样本 X 的概率空间为 $(\mathscr{X}, \mathscr{B}_x, \mathscr{P})$，$\mathscr{P}=\{P_\theta, \theta\in\Theta\}$，$\mu$ 为 $(\mathscr{X}, \mathscr{B}_x)$ 上的一个 σ-有限测度，并设 $\mathscr{P}\ll\mu$，即对每一 $\theta\in\Theta$，存在关于 μ 的概率密度 $p(x, \theta)=dP_\theta(x)/d\mu$. 同以往的约定一样，我们总假定样本空间 $(\mathscr{X}, \mathscr{B}_x)$ 及参数空间 $(\Theta, \mathscr{B}_\Theta)$ 都是欧氏的.

若把 $p(x, \theta)$ 看成是 $\mathscr{X}\times\Theta$ 上的函数，是可测的，我们称之为似然函数. 通常，把密度函数 $p(x, \theta)$ 看成是似然函数. 需强调的是：对每一个固定的 x，把它看成是参数 θ 的函数.

定义 1.1　我们称未知参数 θ 的估计 $\hat{\theta}(x)$ 为其极大似然估计，如果它满足

$$p(x, \hat{\theta})=\sup_{\theta\in\Theta} p(x, \theta) \quad a.s.\mu \tag{1.1}$$

常记为 θ 的 MLE.

又若 $g(\theta)$ 为待估的 k 维向量函数，记 $\Omega=g(\Theta)$，它是 R^k 中的一个子集，称 $g(\theta)$ 的估计 $\hat{g}(x)$ 为它的极大似然估计，如果存在 $\hat{\theta}\in g^{-1}(\hat{g}(x))$，使

$$\sup_{\theta\in\Theta} p(x,\ \theta)=p(x,\ \hat{\theta}) \quad a.s.\mu. \tag{1.2}$$

这个定义是目前一般的数理统计书上所采用的. 因要求 $p(x,\ \theta)$ 在 $\hat{\theta}$ 上有定义, 即要求 $\hat{\theta}\in\Theta$, 从而也要求 $g(\theta)$ 的极大似然估计值需在 Ω 中. 这在实际情况中有时不能满足. 例如, 求二项分布 $B(n,\ p)$中 p 的极大似然估计,若设 $\Theta=(0,1)$, 众所周知, p 的 MLE 为 $\frac{k}{n}$, 其中 k 为试验"成功"次数. 但当 $k=0$ 或 n 时, 则不符合上述定义. 这说明了以上定义是有缺陷的. 为了克服这一缺陷, 我们对定义 1.1 给出如下的补充: 除去满足条件 (1.1)、(1.2)的 $\hat{\theta}$、$\hat{g}$ 分别称为 θ、$g(\theta)$的极大似然估计外,如果 $\hat{\theta}(x)$, $\hat{g}(x)$ 分别满足下面的条件(1.1)′、(1.2)′, 也称为 θ、$g(\theta)$的极大似然估计:

如果 $\hat{\theta}(x)\overline{\in}\Theta$, 但存在一串 $\theta_i\in\Theta$, 使

$$\begin{cases}\lim\limits_{i\to\infty}\theta_i=\hat{\theta},\\ \lim\limits_{i\to\infty}p(x,\ \theta_i)=\sup\limits_{\theta\in\Theta}p(x,\ \theta) \quad a.\ s.\ \mu.\end{cases} \tag{1.1$'$}$$

如果 $\hat{g}(x)\overline{\in}\Omega$, 但存在一串 $\theta_i\in\Theta$, 使

$$\begin{cases}\lim\limits_{i\to\infty}g(\theta_i)=\hat{g}(x),\\ \lim\limits_{i\to\infty}p(x,\ \theta_i)=\sup\limits_{\theta}p(x,\ \theta) \quad a.\ s.\ \mu.\end{cases} \tag{1.2$'$}$$

经过这样补充之后, 在上面说到的 $\frac{k}{n}$ 就确为 p 的 MLE 了.

注 1 在如上的定义中, 我们并没有要求似然函数 $p(x,\ \theta)$一定要关于 θ 可微,当 $p(x,\ \theta)$, $\theta\in\Theta$ 有共同支撑时, 为计算方便起见,有时采用使 $\log p(x,\ \theta)$达到极大的方法来表述 MLE. 又如果它关于 θ 可微, 则往往采用求似然方程 $\frac{\partial}{\partial\theta}\log p(x,\theta)=0$ 的根的办法来求解 θ 的 MLE. 但要指出的是,似然方程的根并非一定是 MLE. 反过来,若似然方程的根不存在,也并不能推定 MLE 不存在.

注 2 极大似然估计有时不唯一.

例 1.1 i)设 $X_1, \cdots, X_n$ 为 $iid.$ 随机变量，服从均匀分布 $R(0, \theta)(0<\theta<\infty)$. 因为

$$dP_\theta(x)=\theta^{-n}I_{[0<\max(x_1,\cdots,x_n)<\theta]}(x)dx,$$

不难看到 $\hat{\theta}(x)=\max\limits_{1\leqslant i\leqslant n}x_i$ 是 θ 的唯一的 MLE.

ii) 设 $X_1, \cdots, X_n$ 为 $iid.$ $X_1\sim R(\theta, \theta+1)(-\infty<\theta<\infty)$.

因为 $dP_\theta(x)=p(x, \theta)dx=I_{[\theta<\min\limits_i x_i<\max\limits_i x_i<\theta+1]}(x)dx$, 于是 $\max\limits_i x_i-1$、$\min\limits_i x_i$ 及它们之间的任一值都是 θ 的极大似然估计，这说明 θ 的 MLE 有无限个.

现在来讨论极大似然估计的不变性原则. 直接从 MLE 的定义出发，立即可得如下定理[1]:

定理 1.1 若 $g(\theta)$ 是在 Θ 上定义的可测向量函数，Θ 为 k 维非退化凸集，$\hat{\theta}(x)$ 为 θ 的 MLE, 且当 $\hat{\theta}(x)\overline{\in}\Theta$ 时，对任意点列 $\{\theta_i\}\subset\Theta$, 若 $\lim\limits_{i\to\infty}\theta_i=\hat{\theta}(x)$, 极限值 $\lim\limits_{i\to\infty}g(\theta_i)$ 存在且相同，并且 $\overline{\in}g(\Theta)$, 那么

$$\hat{g}(x)=\begin{cases}g(\hat{\theta}(x)), & \text{当 } \hat{\theta}(x)\in\Theta;\\ \lim\limits_{\theta_i\to\hat{\theta}(x)}g(\theta_i), & \{\theta_i\}\subset\Theta, \text{ 当 } \hat{\theta}(x)\overline{\in}\Theta.\end{cases}$$

是 $g(\theta)$ 的 MLE.

证 i) 先证 $\hat{g}(x)$ 的可测性: 任取 Θ 的一个内点 θ_0, 定义

$$\hat{g}_i(x)=g(\hat{\theta}(x))I_{[\hat{\theta}(x)\in\Theta]}(x)+g\left(\frac{\theta_0}{i}+\hat{\theta}(x)\left(1-\frac{1}{i}\right)\right)I_{[\hat{\theta}(x)\overline{\in}\Theta]}(x)$$

$(i=1, 2, \cdots)$, 由 Θ 为非退化凸集，$\hat{g}_i(x)$ 有定义，以及 $g(\theta)$、$\hat{\theta}(x)$ 的可测性，故 $\hat{g}_i(x)$ 可测，从而

$$\lim_{i\to\infty}\hat{g}_i(x)=\hat{g}(x)$$

可测，即为 $g(\theta)$ 的一个估计.

ii) $\hat{g}(x)$ 为 MLE, 当 $\hat{\theta}(x)\in\Theta$ 时，因为 $\hat{\theta}(x)$ 为 MLE, 所以有 $\sup\limits_{\theta\in\Theta}p(x, \theta)=p(x, \hat{\theta}(x))$, 但 $\hat{\theta}(x)\in g^{-1}(\hat{g}(x))$, 故此时定义 1.1 满足.

如果 $\hat{\theta}(x)\overline{\in}\Theta$, 由定义 1.1 的补充可知，存在 $\{\theta_i\}\subset\Theta$, 使

$\lim\limits_{i\to\infty}\theta_i=\hat{\theta}(x)$, 且 $\lim\limits_{i\to\infty}p(x,\ \theta_i)=\sup\limits_{\theta}p(x,\ \theta)$, 但又由 $g(\theta)$ 的假设条件及 $\hat{g}(x)$ 定义知,此时有

$$\hat{g}(x)=\lim_{i\to\infty}g(\theta_i).$$

故又满足定义 1.1 的补充. 因此 $\hat{g}(x)$ 为 $g(\theta)$ 的 MLE.

注 3 如果 $\mu(\hat{\theta}(x)\overline{\in}\Theta)=0$, 则定理中关于 Θ 的凸性以及 $g(\theta)$ 在 Θ 的边界上的连续性等假设条件就没有必要了. 此时可令 $\hat{g}(x)=$常数, 当 $\hat{\theta}(x)\overline{\in}\Theta$, $\hat{g}$ 仍为 $g(\theta)$ 的 MLE, 因为在 μ 零集上改变估计值不影响估计的性质.

注 4 即使 θ 的 MLE 不存在,不能断定 $g(\theta)$ 的 MLE 也不存在.

例 1.2 设 $X\sim\mathscr{P}(\lambda)$, $(0<\lambda<\infty)$, 即有概率函数

$$p(x,\ \lambda)=P_\lambda[X=x]=\frac{e^{-\lambda}\lambda^x}{x!}(x=0,\ 1,\ 2,\ \cdots).$$

众所周知, λ 的 MLE 为 $\hat{\lambda}=x$. 但如果 $\theta=\log\lambda$, 因为 $P_\lambda(X=0)=e^{-\lambda}>0$, 而它只在 $\lambda=0$ 或 $\theta=-\infty$ 时, 才达到极大. 这说明 $\theta=\log\lambda$ 的 MLE 不存在.

按照因子判别定理(第一章 § 6), 假若 $T(x)$ 是 θ 的充分统计量, 那么密度可表为 $p(x,\ \theta)=f_\theta(T(x))h(x)$. 可见, θ 的极大似然估计 $\hat{\theta}$ 是充分统计量 $T(x)$ 的函数. 不过, $\hat{\theta}$ 未必为完全充分统计量的函数,因为完全充分统计量不一定存在.

例 1.3 设 $X_1,\ \cdots,\ X_n$ 为 *iid*. 随机变量, $X_1\sim R(\theta,\ 2\theta)$, $0<\theta<\infty$. 则容易求出 θ 的 MLE$\hat{\theta}$ 为 $\min\limits_i x_i$ 或 $\frac{1}{2}\max\limits_i x_i$, 或它们之间任一值. 可见 $\hat{\theta}$ 为完全充分统计量$(\min\limits_i x_i,\ \max\limits_i x_i)$ 的函数. 当然, $\hat{\theta}$ 自身并不是充分统计量,更不是完全充分统计量.

下面举几个例子,说明 UMVUE 与 MLE 的关系.

例 1.4 设 $X_1,\ X_2,\ \cdots,\ X_n$, 为 n 个 *iid*. 随机变量.

i) 当 $X_1\sim R(0,\ \theta)$, $0<\theta<\infty$, 则 θ 的 UMVUE 为:

$$\left(1+\frac{1}{n}\right)\max_{1\le i\le n}x_i,\quad E\left[\left(1+\frac{1}{n}\right)\max_i X_i-\theta\right]^2=\frac{\theta^2}{n(n+2)};$$

θ 的 MLE 为 $\quad \max_i x_i,\ E\,[\max_i X_i-\theta]^2=\dfrac{2\theta^2}{(n+1)(n+2)}$;

θ 的最小平方损失风险估计为:

$$\frac{n+2}{n+1}\max_i x_i,\ E\left[\frac{n+2}{n+1}\max_i X_i-\theta\right]^2=\frac{\theta^2}{(n+1)^2}.$$

ii) 当 $X_1\sim N(\mu,\ \sigma^2)$, $-\infty<\mu<\infty$, $0<\sigma^2<\infty$, 则 σ^2 的 UMVUE 为:

$$\frac{1}{n-1}\sum_{i=1}^n(x_i-\bar{x})^2,\quad E\left[\frac{1}{n-1}\sum_{i=1}^n(x_i-\bar{x})^2-\sigma^2\right]^2=\frac{2\sigma^4}{n-1};$$

σ^2 的 MLE 为

$$\frac{1}{n}\sum_{i=1}^n(x_i-\bar{x})^2,\ E\left[\frac{1}{n}\sum_{i=1}^n(x_i-\bar{x})^2-\sigma^2\right]^2=\frac{2n-1}{n^2}\sigma^4;$$

σ^2 的最小平方损失风险估计为:

$$\frac{1}{n+1}\sum_{i=1}^n(x_i-\bar{x})^2,$$

$$E\left[\frac{1}{n+1}\sum_{i=1}^n(x_i-\bar{x})^2-\sigma^2\right]^2=\frac{2}{n+1}\sigma^4.$$

iii) 当 $X_1\sim B(1,\ p)$, $0<p<1$, 记 $q=1-p$, 则 $\theta=pq$ 的 UMVUE 为:

$$\frac{\left(\sum_1^n x_i\right)\left(n-\sum_{i=1}^n x_i\right)}{n(n-1)},\quad E\left[\theta-\frac{\left(\sum_{i=1}^n x_i\right)\left(n-\sum_{i=1}^n x_i\right)}{n(n-1)}\right]^2$$

$$=\frac{\theta}{n}-\frac{2(2n-3)}{n(n-1)}\theta;$$

MLE 为:

$$\frac{1}{n^2}\left(\sum_{i=1}^n x_i\right)\left(n-\sum_{i=1}^n x_i\right),$$

$$E\left[\theta-\frac{1}{n^2}\left(\sum_1^n x_i\right)\left(n-\sum_1^n x_i\right)\right]^2$$

$$=\frac{(n-1)^2}{n^3}\theta-\frac{4n^2-9n+6}{n^3}\theta^2.$$

此例说明，就平方损失而言，MLE 与 UMVUE 相比，随分布不同而各有千秋，都不如最小平方损失估计.

§2 指数型分布族参数的 MLE

一、指数型分布族的某些进一步的性质

我们在第一章§2及第二章§4中已介绍了它的许多优良性质. 为研究指数族的 MLE 问题, 我们需要对它的性质作进一步补充. 关于这方面的研究,有兴趣的读者还可参考文献 [2], [3], [4], [5] i).

我们考虑标准化的指数族

$$d P_{\theta}(x)=\beta(\theta) e^{\theta' x} d\mu \quad (\theta \in \Omega), \tag{2.1}$$

其中, 我们设样本空间$(\mathscr{X}, \mathscr{B}_x)$为 k 维欧氏的, μ 为其上 σ-有限测度, $\beta(\theta)=\left[\int_{x} e^{\theta' x} d\mu(x)\right]^{-1}$, $\Omega=\{\theta: 0<\beta(\theta)<\infty\}$ 为自然参数空间, $\Omega \subset R^{k}$.

首先来明确参数 θ 与 x 的有意义的值域. 我们已知道 Ω 是 R^{k} 中的凸集. 记

$\Omega^{0}=\Omega$ 的内点集合,即 Ω 的内部;

$m(\theta)=E_{\theta}(X)=\beta(\theta) \int_{x} x e^{\theta' x} d\mu$;

$M=\{m(\theta): -\infty<m(\theta)<\infty, \theta \in \Omega\}$;

$D=\{x_{0}$: 对任意 $\varepsilon>0, \mu(\|x-x_{0}\|<\varepsilon)>0, x_{0} \in X\}$;

$C=D$ 的凸闭包,即包含 D 的最小闭凸集,

$\bar{C}=C$ 的边界;

$C^{0}=C$ 的内点的集合,即 C 的内部.

由指数族的性质(第一章定理 2.2) 及上面记号的定义,不难推得:

i) 当 $\theta \in \Omega^{0}$ 时, 则 $|m(\theta)|<\infty$. 记 $M^{0}=M$ 的内部, 则 $M^{0}=\{m(\theta): \theta \in \Omega^{0}\}$;

ii) $\mu(\mathscr{X}-D)=\mu(\mathscr{X}-C)=0$, 从而对任意的 $\theta \in \Omega, P_{\theta}(\mathscr{X}-C)=0$. 这说明 $\mathscr{X}$ 与 C 等价.

我们可以看到 θ 的函数 $m(\theta)$ 及 θ 本身有它特殊的意义. 因

此在这节我们仅讨论 θ 和 $m(\theta)$ 的 MLE. 由于 MLE 有不变性，似乎只要弄清 θ 的 MLE 就可以了. 但事实并不完全如此，因为以后我们会看到，有时即使 θ 的 MLE 不存在，但是 $m(\theta)$ 的 MLE 依然存在.

以下我们皆假设 C^0 非空，亦即表示 X 为非退化的 k 维随机变量. 其次我们假设 Ω^0 非空，如果 Ω^0 是空集，这表示 Ω 实质是 $R^h(h<k)$ 内的凸集，此时我们可以降维，把指数族看成是 h 维分布族就可以了. 因此,这两个基本假设是不失一般性的假设.

如果对指数族 (2.1) 进行 n 次独立抽样 $X_1, \cdots, X_n$，注意 $\overline{X}=\frac{1}{n}\sum_1^n X_i$ 是完全充分统计量,且 $\overline{X}$ 的分布仍为指数族,自然,参数空间 Ω 和样本空间均不变. 因此我们只需考虑 $n=1$ 的情况. 但需要指出的是:若要细究，$\overline{X}$ 与 X 的值域 C 虽然一样，但概率结构却不同了.

引理2.1 i) $-\log\beta(\theta)$ 在 Ω 上是严格凸函数; ii) $M\subset C^0$; iii) $M_1=\{m(\theta),\ \theta\in\Omega^0\}$ 与 Ω^0 一一对应.

证 i)在 Ω 中任取两点 $\theta_0\neq\theta_1$ 及 $0<p<1$，利用 Hölder 不等式,有

$$\begin{aligned}-\log\beta(p\theta_0+(1-p)\theta_1)&=\log\int e^{(p\theta_0+(1-p)\theta_1)'x}d\mu\\ &\leqslant\log\left\{\left[\int e^{\theta_0'x}\,d\mu\right]^p\left[\int e^{\theta_1'x}d\mu\right]^{1-p}\right\}\\ &=p\log\int e^{\theta_0'x}d\mu+(1-p)\log\int e^{\theta_1'x}\,d\mu\\ &=-p\log\beta(\theta_0)-(1-p)\log\beta(\theta_1).\end{aligned}\tag{2.2}$$

而且使等号成立 $\Leftrightarrow$ 存在常数 c，使

$$\begin{aligned}e^{(1-p)\theta_1'x}=ce^{p\theta_0'x\left(\frac{1-p}{p}\right)}=ce^{(1-p)\theta_0'x}\quad a.\ s.\ \mu\Leftrightarrow(1-p)\theta_1'x\\ =(1-p)\theta_0'x,\ a.\ s.\ \mu.\end{aligned}$$

但因 $\theta_1\neq\theta_0$，这不可能成立，故上面只能取“$<$”号，这就证明了 $-\log\beta(\theta)$ 为严凸函数.

ii) 若对某个 $\theta\in\Omega$，$m(\theta)$ 有限，但 $m(\theta)\overline{\in}C^0$，因为 C 是凸

集，故由第一章§3的定理3.1知：存在支撑平面$a'(x-m(\theta))=0$，使当$x\in C$时，有$a'(x-m(\theta))\geqslant 0$，即$P_\theta\{a'(X-m(\theta))\geqslant 0\}=1$。但$E_\theta a'(X-m(\theta))=a'(m(\theta)-m(\theta))=0$，从而$P_\theta\{a'(X-m(\theta))=0\}=1$。这说明$C$几乎全是边界，这与$C^0$为非空的假设相矛盾。

iii）设$\theta_1\neq\theta_2(\theta_1、\theta_2\in\Omega)$。由$e^x$是严凸的，$X$是非退化的，利用Jensen不等式，得

$$1=\frac{\beta(\theta_2)}{\beta(\theta_1)}E_{\theta_1}\left[e^{\theta_2'X-\theta_1'X}\right]>\frac{\beta(\theta_2)}{\beta(\theta_1)}e^{(\theta_2-\theta_1)'m(\theta_1)}.$$

同理$\frac{\beta(\theta_1)}{\beta(\theta_2)}e^{(\theta_1-\theta_2)'m(\theta_2)}<1$。如果$m(\theta_1)=m(\theta_2)$，显然与上面的两个不等式相矛盾，故iii)的结论成立。

引理2.2 i）$x_0\in C^0\Leftrightarrow$对任意$\|a\|=1$的向量$a\in R^k$，皆有$\mu(x:a'(x-x_0)>0)>0$；

ii）若$x_0\in C^0$，对任意$\{\theta_i\}\subset\Omega$，如果$\lim\limits_{i\to\infty}\theta_i=\theta_0\overline{\in}\Omega$或者$\lim\limits_{i\to\infty}\|\theta_i\|=\infty$，皆有

$$\lim_{i\to\infty}p(x_0,\ \theta_i)=0,\quad 其中\quad p(x,\ \theta)=\beta(\theta)e^{\theta'x}.\tag{2.3}$$

证 “i）$\Rightarrow$”：对任意$a_0\in R^k$，$\|a_0\|=1$。因为x_0是C的内点，故$C^0\cap\{x:a_0'(x-x_0)>0\}$是非空的，且为凸集．由于$C$是包含$D$的最小凸闭集以及由$D$的定义，有

$$\mu\{x:a_0'(x-x_0)>0\}>0.$$

“i）$\Leftarrow$”：假设$x_0\overline{\in}C^0$，则x_0或为C的边界点或外点，因此由C的凸性，存在$\|a_0\|=1$，$a_0\in R^k$，使当$x\in C$时，$a_0'(x-x_0)\leqslant 0$。故$\mu(x:a_0'(x-x_0)>0)\leqslant\mu(\mathscr{X}-C)=0$。从而i)的充分性得证。

ii）设x_0是C的内点，则一定存在充分小的$\varepsilon>0$，使

$$\inf_{\|a\|=1}\mu\{x:a'(x-x_0)>\varepsilon\}=\delta(\varepsilon)>0.\tag{2.4}$$

事实上，若取$\varepsilon>0$，使x_0的2ε-邻域$V_{2\varepsilon}(x_0)=\{x:\|x-x_0\|<2\varepsilon\}\subset C^0$。存在$\{a_i\}\subset R^k$，$\|a_i\|=1(i=1,\ 2,\ \cdots)$，使

$$\delta(\varepsilon)=\lim_{i\to\infty}\mu\{x:a_i'(x-x_0)>\varepsilon\}.$$

因 $\{a_i\}$ 有界，不妨设 $\lim\limits_{i\to\infty} a_i = a_0 \in R^k$. 于是由测度的连续性知：有 $\delta(\varepsilon) = \mu\{x: a_0'(x-x_0) > \varepsilon\}$，但 x_0 的 ε-邻域 $V_\varepsilon(x_0)$ 的边界上一定存在一个点 x_1，使

$$\{x: a_0'(x-x_0) > \varepsilon\} = \{x: a_0'(x-x_1) > 0\}.$$

注意：x_1 为 C 的内点，由 i)知 $\delta(\varepsilon) = \mu\{x: a_0'(x-x_1)\} > 0$. 于是(2.4)得证．从而由 $p(x, \theta)$ 的定义及(2.4)式，有

$$\limsup_{i\to\infty} p(x_0, \theta_i) = \limsup_{i\to\infty} \frac{e^{\theta_i' x_0}}{\int e^{\theta_i' x} d\mu} = \frac{1}{\liminf\limits_{i\to\infty} \int e^{\theta_i'(x-x_0)} d\mu}$$

$$\leqslant 1 \Big/ \liminf_{i\to\infty} \int_{\{\theta_i'(x-x_0)/\|\theta_i\| > \varepsilon\}} e^{\theta_i'(x-x_0)} d\mu$$

$$\leqslant \lim_{i\to\infty} e^{-\|\theta_i\|\varepsilon} \frac{1}{\delta(\varepsilon)} = 0 \quad (\text{当 } \|\theta_i\| \to \infty). \tag{2.5}$$

另一方面，如果 $\lim\limits_{i\to\infty} \theta_i = \theta_0 \overline{\in} \Omega$，由 Ω 的定义知，

$$\beta(\theta_0) = \left[\int e^{\theta_0' x}\, d\mu\right]^{-1} = 0,$$

故有

$$\limsup_{i\to\infty} p(x_0, \theta_i) = \lim_{i\to\infty} e^{\theta_i' x_0} \Big/ \liminf_{i\to\infty} \int e^{\theta_i' x}\, d\mu$$

$$\leqslant e^{\theta_0' x_0} \Big/ \int \lim_{i\to\infty} e^{\theta_i' x}\, d\mu = 0.$$

至此，ii)得证，引理全部证毕．

二、指数族参数的 MLE

利用上面的结果，我们可以证明如下 θ 的 MLE 存在唯一性定理．

定理 2.1 若 $\mu(C - C^0) = \mu(\bar{C}) = 0$，且 $\Omega = \Omega^0$. 则有

(i) 似然方程

$$x = m(\theta) \tag{2.6}$$

当 $x \in C^0$ 时，有唯一的根 $\theta(x)$ 存在；

$$\text{(ii)} \quad \hat{\theta}(x) = \begin{cases} \theta(x), & \text{当 } x \in C^0; \\ \theta_0, & \text{当 } x \overline{\in} C^0, \end{cases} \tag{2.7}$$

是 θ 的基本唯一的 MLE. (2.7)式中的 θ_0 是 Ω 中的任一定值.

证 分以下几步来证:

(a) 因为 $\mu(\bar{C})=0$, 故当 $x\in\bar{C}^0$ 时, 由 MLE 的定义, 此时 θ 的 MLE 在 x 点可取任意值 $\theta_0\in\Omega$.

(b) 设 $\Omega=\Omega^0$. 当 $x\in C^0$ 时, 由引理 2.2 知, 此时似然函数 $p(x,\theta)$ 只能在 Ω 中达到极大. 但又由引理 2.1 知, $-\log\beta(\theta)$ 是 θ 的严格凸函数, 从而 $\theta'x+\log\beta(\theta)$ 为 θ 的严格凹函数, 因此在 Ω 上存在唯一的极大值点 $\theta(x)$. 这就证明了 θ 的 MLE 存在且基本唯一. 但 $\theta(x)\in\Omega=\Omega^0$, 故必满足似然方程

$$x=-\frac{\partial}{\partial\theta}\log\beta(\theta)=m(\theta).$$

因为 $\theta_1\neq\theta_2\Rightarrow m(\theta_1)\neq m(\theta_2)$, 故此似然方程仅有唯一的根 $\theta(x)$. 定理证毕.

注 定理的结论对于 $\Omega-\Omega^0\neq\emptyset$, 当 $\theta\in\Omega-\Omega^0$, $E_\theta\|X\|=\infty$ 的情况下, 仍然成立(参看[2]).

定理 2.2 如果 $k=1$, 即在单参数的情况, 则 $m(\theta)$ 的 MLE 存在, 而且基本唯一($a.s.\ \mu$)地可表示为

$$\hat{m}(x)=\begin{cases} x, & \text{当 } x\in M;\\ a, & \text{当 } x\leqslant a, \text{ 假定 } M \text{ 存在有限左端点 } a;\\ b, & \text{当 } x\geqslant b, \text{ 假定 } M \text{ 存在有限右端点 } b.\end{cases}\tag{2.8}$$

证 情况 1: $\mu(\bar{C})=0$, 在定理 2.1 已讨论了 θ 的 MLE $\hat{\theta}(x)$ 基本唯一. 当 $x\in C^0$ 时, $\hat{\theta}(x)$ 为似然方程 $x=-\dfrac{\beta'(\theta)}{\beta(\theta)}=m(\theta)$ 的根. 当 $x\in\bar{C}$, 可以任意取定一个值. 由于 $m(\theta)$ 为单调严增函数, 满足 MLE 不变性定理的条件. 故当 $x\in C^0$, $m(\theta)$ 的 MLE $\hat{m}(x)=m(\hat{\theta}(x))=x$. 当 $x\in\bar{C}$ 时, 可按(2.8)式取值. 由 $\mu(\bar{C})=0$, 不论 $\hat{m}(x)$ 在 $\bar{C}$ 上如何取值, 都不影响其 MLE 性质. 至此, 在情况 1 下定理证毕.

情况 2: $\mu(\bar{C})>0$. 记 C 的端点为 x_m 和 x_M, 显然, 其中至少有一个有限, 且使在该点 μ 的测度为正. 不妨设 $\mu(\{x_m\})>0$.

$$p(x_m,\ \theta)=\frac{\theta x_m}{\int e^{\theta x}\,d\mu}=1\Big/\int_C e^{\theta(x-x_m)}\,d\mu.$$

因为 $x\geqslant x_m$, 当 $x\in C$, 故 $\int_C e^{\theta(x-x_m)}\,d\mu$ 是 θ 的严格上升函数(因为 C^0 非空), 从而 $p(x_m,\theta)$ 为 θ 的严格下降函数. 而当 $\theta<\theta_0\in\Omega^0$ 时,

$$\begin{aligned}\int e^{\theta x}\,d\mu&=\int_{x_m\leqslant x<0}e^{\theta x}\,d\mu+\int_{x\geqslant 0}e^{\theta x}\,d\mu\\&\leqslant e^{(\theta-\theta_0)x_m}\int_{x_m\leqslant x<0}e^{\theta_0 x}\,d\mu+\int_{x\geqslant 0}e^{\theta_0 x}\,d\mu<\infty.\end{aligned}$$

即有 $0<\beta(\theta)<\infty$, 从而 $\theta\in\Omega$, 这说明 Ω 无有限左端点. 此时, $p(x_m,\ \theta)$ 只能在 $\theta\to-\infty$ 时才能达到极大. 但另一方面,

$$\begin{aligned}\lim_{\theta\to-\infty}m(\theta)&=\lim_{\theta\to-\infty}\frac{\int_C xe^{\theta(x-x_m)}\,d\mu}{\int_C e^{\theta(x-x_m)}\,d\mu}=\frac{x_m\mu(\{x_m\})}{\mu(\{x_m\})}\\&=x_m=M\ \text{的左端点}\ a.\end{aligned}\tag{2.9}$$

同样, 对 $\mu(\{x_M\})>0$ 的情况, 可以证明 $p(x_M,\ \theta)$ 为 θ 的严格上升函数, Ω 无有限右端点. 故它仅当 $\theta\to\infty$ 时达到极大, 而且

$$\lim_{\theta\to\infty}m(\theta)=x_M=M\ \text{的右端点}\ b.\tag{2.10}$$

那么由 MLE 的定义可推得

$$\hat{m}(x_m)=a,\quad \text{当}\ \mu(\{x_m\})>0\ \text{时};\tag{2.11}$$

$$\hat{m}(x_M)=b,\quad \text{当}\ \mu(\{x_M\})>0\ \text{时}.\tag{2.12}$$

当 $x\in C^0$ 时, 情况完全与情况 1 的分析相同. 但应当指出, 如果 $\mu(\{x_m\})>0$, 就不会出现 $x\in C$, $x<a$ 的情况, 若 $\mu(\{x_M\})>0$, 就不会出现 $x\in C$, $x>b$ 的情况.

根据情况 1、2 的分析, 就证明了 (2.8) 式所定义的 $\hat{m}(x)$ 是 $m(\theta)$ 的基本唯一的 MLE. 定理得证.

注 3 在定理的证明中情况 2 的证明过程中已说明了当 $\mu(\bar{C})>0$ 时, 由于 $p(x_m,\ \theta)$ 或 $p(x_M,\ \theta)$ 只能在无穷远处才能达到极大, 说明此时 θ 的 MLE 不存在. 但是我们已看到 $m(\theta)$ 的 MLE 却依然存在. 此事就说明有了 θ 的 MLE 存在性的定理后, 仍有

必要讨论 $m(\theta)$ 的 MLE 存在性问题. 例如二项分布及 Poisson 分布都属于此类情况.

例 2.1 若 $X_1, \cdots, X_n$ 为 iid. 随机向量(维数 $p\leqslant n$), $X_1\sim N(\mu, A)$, 此处 $A>0$ 为 p 阶方阵, μ 为 p 维向量. 则 $X=(X_1, \cdots, X_n)$ 的联合概率密度是

$$(2\pi)^{-\frac{pn}{2}}(\det A)^{-n/2}\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)'A^{-1}(x_i-\mu)\right\}$$

$$=(2\pi)^{-\frac{pn}{2}}(\det A)^{\frac{-n}{2}}\exp\left\{-\frac{1}{2}\operatorname{tr}A^{-1}\Sigma x_i x_i'\right.$$

$$\left.+\sum_{i=1}^{n}x_i'A^{-1}\mu-\frac{n}{2}\mu'A^{-1}\mu\right\}.$$

它是 $p+\frac{p(p+1)}{2}=\frac{p(p+3)}{2}$ 维指数分布族. 记

$$T=\left(\sum_{i=1}^{n}x_i, \sum_{i=1}^{n}x_i x_i'\right)=(T_1, T_2);$$

$$\theta=\left(A^{-1}\mu, -\frac{1}{2}A^{-1}\right);$$

$$\Omega=\left(A^{-1}\mu\in R^p, -\frac{1}{2}A^{-1}<0\right)=\Omega^0;$$

$$E_\theta T=(n\mu, nA+n\mu\mu')\triangleq m(\theta);$$

C 为 T 的值域, 为 $\{T_1\in R^p, T_2>0\}$, 它是 $R^{\frac{p(p+3)}{2}}$ 中的一个凸集. 由于分布的连续性及 $L(C-C^0)=0$, 根据定理 2.1, 似然方程为

$$n\mu=\sum_{i=1}^{n}x_i, \quad n(A+\mu\mu')=T_2=\sum_{i=1}^{n}x_i x_i'.$$

它有唯一解

$$\hat{\mu}=\frac{1}{n}\sum_{i=1}^{n}x_i, \quad \hat{A}=\frac{1}{n}\sum_{i=1}^{n}x_i x_i'-\hat{\mu}\hat{\mu}'=\frac{1}{n}\sum_{i=1}^{n}(x_i-\hat{\mu})(x_i-\hat{\mu})',$$

这就是 μ、A 的 MLE. 至于 θ 的 MLE, 是

$$\hat{\theta}=\left(\hat{A}^{-1}\hat{\mu}, -\frac{1}{2}\hat{A}^{-1}\right).$$

由极大似然估计的不变性, 我们还可得到

$$\Phi(B)=(2\pi)^{-\frac{p}{2}}(\det A)^{-\frac{1}{2}}\int_B e^{-\frac{1}{2}(x-\mu)'A^{-1}(x-\mu)}dx$$

的 MLE 为 $\hat{\Phi}(x)=(2\pi)^{-\frac{p}{2}}(\det\hat{A})^{-\frac{1}{2}}\int_B e^{-\frac{1}{2}(x-\hat{\mu})'\hat{A}^{-1}(x-\hat{\mu})}dx.$

例 2.2 设 $X_1, \cdots, X_n$ 为 $iid.$ 随机变量，具有关于 μ 的概率密度

$$p(x, \theta)=\beta(\theta)e^{\theta x}.$$

此处 $$d\mu=\begin{cases}x^{-r}dx, & 当\ 1<x<\infty \\ \mu\{1\}>0 & \end{cases}\quad (r\geqslant 0).$$

$$\Omega=\begin{cases}\{\theta:0<\beta(\theta)<\infty\}=\{\theta: -\infty<\theta<0\} & (当\ r\leqslant 1); \\ \{\theta: -\infty<\theta\leqslant 0\} & (当\ r>1).\end{cases}$$

当 $0\leqslant r\leqslant 1$ 时，$\Omega=\Omega^0$. $m(\theta)$ 的 MLE 为 $\hat{m}=\bar{x}$. 当 $1<r\leqslant 2$ 时，$E_0|\bar{X}|=\infty$. 故 $m(\theta)$ 的值域 M 为 $(1, \infty)$. 因而 $m(\theta)$ 的 MLE 也为 $\hat{m}=\bar{X}$. 但当 $r>2$ 时，

$$E_0\bar{X}=E_0X_1=\frac{1-r}{2-r}\frac{(r-2)\mu\{1\}+1}{(r-1)\mu\{1\}+1}\triangleq b,$$

于是 $M=(1, b]$. 此时有

$$\hat{m}(x)=\begin{cases}\bar{x}, & 当\quad 1\leqslant\bar{x}\leqslant b; \\ b, & 当\quad \bar{x}>b.\end{cases}$$

最后，我们指出，若指数族给定的原始的形式

$$p(x, \theta)d\mu=\beta(\theta)\exp\left\{\sum_{i=1}^{k}\theta_i(\varphi_1, \cdots, \varphi_k)T_i(x)\right\}d\mu$$

当 $\theta=(\theta_1, \cdots, \theta_k)'$ 的自然区域 Ω 为开集时，θ 的 MLE 也存在. 我们可以推得 θ 的 MLE 是满足方程 $E_\theta T=T$ 的根，而 $m(\theta)=E_\theta T$ 的 MLE 为 $\hat{m}=T$. 若 $\varphi=(\varphi_1, \cdots, \varphi_k)'$ 可表为 θ 的函数，且 $H=\left(\frac{\partial\varphi_i}{\partial\theta_j}\right)$ 非奇异，由不变性原则知：$\hat{\varphi}=\varphi(\hat{\theta})$ 是 φ 的 MLE. 众所周知，Fisher 信息阵

$$I(\theta)=E_\theta\left[\left(\frac{\partial\log p(x, \theta)}{\partial\theta}\right)\left(\frac{\partial\log p(x, \theta)}{\partial\theta}\right)'\right]=\mathrm{cov}_\theta T$$

$$=-\left(\frac{\partial^2\log\beta(\theta)}{\partial\theta_i\partial\theta_j}\right). \tag{2.13}$$

而关于 φ 的 Fisher 信息阵则是

$$I(\varphi)=(H^{-1})'I(\theta(\varphi))(H^{-1}). \tag{2.14}$$

这样，我们初步讨论了指数分布族参数 θ 以及均值 $m(\theta)$ 的极大似然估计，给出了其 MLE 存在、唯一的充分条件和计算表达式. 细心的读者会发现，这些结果仍可作更深入的讨论. 事实上，θ 的 MLE 存在唯一的充要条件已在 [5] i) 中给出，有兴趣的读者可以参阅.

§3 极大似然估计的相合性

一、MLE 强相合性的一般定理

设 $X_1, \cdots, X_n, \cdots$ 为 $iid.$ 随机变量序列，它们具有关于 σ-有限测度 μ 的概率密度 $p(x, \theta)=dp_\theta(x)/d\mu$, $\theta\in\Theta$, 并设 Θ 为 R^p 中的一个开域. $\hat{\theta}_n(x_1, \cdots, x_n)$ 为 θ 的基于 $x_1, \cdots, x_n$ 的 MLE· 关于 MLE$\hat{\theta}_n$ 的强相合性问题，一些著名的数理统计学家一直在研究. 例如 Wald[6], Le Cam[7], Huber[8], Bahadur[9] 等都作过贡献. 我们在这里采用 M. D. Perlman[10] 工作中比较具体而又有广泛应用的形式，其具体证明由卢昆亮所给出.

记 $\overline{\Theta}$ 为 Θ 的闭包，必要时可使其包括无穷远点，使 $\overline{\Theta}$ 构成紧集. 对于 $n>m\geqslant 0$, θ、$\varphi\in\Theta$, 记 $\mathscr{P}=\{P_\theta, \theta\in\Theta\}$; 并记

$$l_{m,n}(\theta, \varphi)=\begin{cases}\log\prod\limits_{i=m+1}^{n}\dfrac{p(x_i, \theta)}{p(x_i, \varphi)}, & \text{当} \prod\limits_{i=m+1}^{n}p(x_i, \theta)>0;\\ 0, & \text{其它}.\end{cases} \tag{3.1}$$

$$I_{m,n}(\theta, \varphi)=E_\theta\{l_{m,n}(\theta, \varphi)\}; \tag{3.2}$$

$$l_{m,n}(\theta, N_\varphi)=\inf_{\xi\in N_\varphi\cap\Theta} l_{m,n}(\theta, \xi). \tag{3.3}$$

此处 N_φ 为 φ 的邻域.

$$I_{m,n}(\theta, N_\varphi)=E_\theta\{l_{m,n}(\theta, N_\varphi)\}. \tag{3.4}$$

易见 $I_{n,n+m}(\theta, \varphi)=I_{0,m}(\theta, \varphi)$, $I_{n,n+m}(\theta, N_\varphi)=I_{0,m}(\theta, N_\varphi)$, 故可把它们分别简记为 $I_m(\theta, \varphi)$, $I_m(\theta, N_\varphi)$.

定理3.1 假设

1）对任意$\theta\neq\varphi(\in\boldsymbol{\Theta})$，$P_\theta(p(X,\varphi)\neq p(X,\theta))>0$；

2）对任意$x\in\mathscr{X}$，$p(x,\varphi)$是$\boldsymbol{\Theta}$上的上半连续函数. 对任意$\varphi\in\overline{\boldsymbol{\Theta}}-\boldsymbol{\Theta}$，存在$m\geqslant1$，有$\limsup\limits_{\Theta\ni\xi\to\varphi}\prod\limits_{i=1}^{m}p(x_i,\xi)<\infty$，$a.s.\ \mathscr{P}^m$，且对一切$\theta\in\boldsymbol{\Theta}$，$P_\theta\left\{\limsup\limits_{\Theta\ni\xi\to\varphi}\prod\limits_{i=1}^{m}p(X_i,\xi)=0\right\}>0$.

3）存在$m\geqslant1$，使对一切$\theta\in\boldsymbol{\Theta}$，$\varphi\in\overline{\boldsymbol{\Theta}}$，$\varphi\neq\theta$，皆存在开邻域$N_\varphi$，使$I_m(\theta,N_\varphi)>-\infty$.

若$\hat{\theta}_n=\hat{\theta}_n(x_1,\cdots,x_n)$为$\theta$的MLE，则对一切$\theta\in\boldsymbol{\Theta}$，有

$$P_\theta(\lim_{n\to\infty}\hat{\theta}_n=\theta)=1.$$

证 分以下几步来证明此定理：

i）对任意的$\theta\in\boldsymbol{\Theta}$，$\varphi\in\overline{\boldsymbol{\Theta}}$，$\varphi\neq\theta$，则$I_m(\theta,\varphi)$存在，而且$I_m(\theta,\varphi)>0$.

由条件3)知，当$\varphi\in\boldsymbol{\Theta}$时，$I_m(\theta,\varphi)\geqslant I_m(\theta,N_\varphi)>-\infty$. 再由条件1)和Jenson不等式，就有$I_m(\theta,\varphi)>0$.

当$\varphi\in\overline{\boldsymbol{\Theta}}-\boldsymbol{\Theta}$时，定义$p(x,\varphi)=\limsup\limits_{\Theta\ni\xi\to\varphi}p(x,\xi)$，$a.s.\ \mathscr{P}$. 由条件2)知，$p(x,\varphi)<\infty$，$a.s.\ \mathscr{P}$. 从而可以定义$l_{m,n}(\theta,\varphi)$，$I_{m,n}(\theta,\varphi)$. 由条件3)，有

$$I_m(\theta,\varphi)\geqslant E_\theta(\inf_{\xi\in N_\varphi\cap\Theta}l_{0,m}(\theta,\xi))=I_m(\theta,N_\varphi)>-\infty. \tag{3.5}$$

因为

$$\begin{aligned}I_m(\theta,\varphi)=&E_\theta(l_{0,m}(\theta,\varphi)I_{\{\prod_{i=1}^{m}p(x_i,\varphi)>0\}}(x_1,\cdots,x_m))\\&+E_\theta(l_{0,m}(\theta,\varphi)I_{\{\prod_{i=1}^{m}p(x_i,\varphi)=0\}}(x_1,\cdots,x_m)).\end{aligned} \tag{3.6}$$

由条件2)知，上式右端第二项等于∞，而第一项大于$-\infty$，从而$I_m(\theta,\varphi)=\infty$.

ii）对任意的$\theta\in\boldsymbol{\Theta}$，$\varphi\in\overline{\boldsymbol{\Theta}}$，$\varphi\neq\theta$，存在$\varphi$的邻域$N_\varphi$，使得$I_m(\theta,N_\varphi)>0$.

根据条件3)，存在φ的邻域$N_\varphi^{(0)}$，使得$I_m(\theta,N_\varphi^{(0)})>-\infty$，

设 $N_\varphi^{(i)}$ $(i=1, 2, \cdots)$是一串 φ 的邻域，满足 $N_\varphi^{(0)}\supset N_\varphi^{(1)}\supset N_\varphi^{(2)}\supset\cdots$，并且当 φ 为有限点时，$\lim\limits_{i\to\infty}N_\varphi^{(i)}=\{\varphi\}$，当 φ 为无穷远点时，此时 N_φ 表示有界闭集的余集，要求 $\lim\limits_{i\to\infty}\inf\limits_{\xi\in N_\varphi^{(i)}}\|\xi\|=\infty$. 易见

$$-\infty<I_m(\theta,\ N_\varphi^{(0)})\leqslant I_m(\theta,\ N_\varphi^{(1)})\leqslant\cdots.$$

应用 $p(x,\ \varphi)$关于 φ 的上半连续性，根据 i)和单调收敛定理，有

$$\lim_{i\to\infty} I_m(\theta,\ N_\varphi^{(i)})\geqslant E_\theta(l_{0,m}(\theta,\ \varphi))>0. \tag{3.7}$$

故存在 i_0，使得 $I_m(\theta,\ N_\varphi^{(i)})>0$，取 $N_\varphi=N_\varphi^{(i_0)}$，结论 ii)得证.

iii) 对任意的 $\varepsilon>0$，令 $N_\theta=\{\xi:\ \|\xi-\theta\|<\varepsilon\}$，记 $N_\theta^c=\overline{\Theta}-N_\theta$. 易见 N_θ^c 为紧集，由 Heine-Borel 有限覆盖定理以及(ii)，存在有限个开邻域 N_{φ_1}, N_{φ_2}, $\cdots$, N_{φ_k} (其中包括一个 ∞ 的邻域 N_∞)，使得 $\bigcup\limits_{i=1}^{k} N_{\varphi_i}\supset N_\theta^c$，且 $I_m(\theta,\ N_{\varphi_i})>0$ $(i=1,\ \cdots,\ k)$.

因为 $I_m(\theta,\ N_{\varphi_i})>0$ $(i=1,\ \cdots,\ k)$，k 有限，则存在充分大的正数 M，使得如下定义的

$$l_{0,m}^{(M)}(\theta,\ N_{\varphi_i})=\begin{cases} l_{0,m}(\theta,\ N_{\varphi_i}), & \text{当 } l_{0,m}(\theta,\ N_{\varphi_i})\leqslant M; \\ M, & \text{当 } l_{0,m}(\theta,\ N_{\varphi_i})>M. \end{cases} \tag{3.8}$$

有 $$M\geqslant E_\theta(l_{0,m}^{M}(\theta,\ N_{\varphi_i}))>0 \quad (i=1,\ \cdots,\ k).$$

对于任意的 n, $n=rm+h$, $0\leqslant h<m$, 注意到

$$l_{jm+h,(j+1)m+h}(\theta,\ N_{\varphi_i}) \quad (j=0,\ \cdots,\ r-1)$$

是 iid. 随机变量，因此，我们若如同 $l_{0,m}^{(M)}(\theta,\ N_{\varphi_i})$一样定义 $l_{jm+h,(j+1)m+h}^{(M)}(\theta, N_{\varphi_i})$, $l_{0,h}^{(M)}(\theta, N_{\varphi_i})$，并把前者简记为 $l_{jm+h}^{(M)}(\theta,\ N_{\varphi_i})$，那么由柯尔莫哥洛夫强大数定律，有

$$\lim_{r\to\infty}\frac{1}{r}\sum_{j=0}^{r-1} l_{jm+h}^{(M)}(\theta,\ N_{\varphi_i})=E_\theta l_{0,m}^{(M)}(\theta,\ N_{\varphi_i}), \quad a.s.\ P_\theta. \tag{3.9}$$

由于 $0\leqslant h<m$, 故

$$\lim_{r\to\infty}\frac{1}{r}\left\{\sum_{j=0}^{r-1} l_{jm+h}^{(M)}(\theta,\ N_{\varphi_i})+l_{0,h}^{(M)}(\theta,\ N_{\varphi_i})\right\}=E_\theta l_{0,m}^{(M)}(\theta,\ N_{\varphi_i}). \tag{3.10}$$

注意：$E_\theta l_{0,m}^{(M)}(\theta, N_{\varphi_i})>0$，于是

$$\left\{\sum_{j=0}^{r-1} l_{jm+h}^{(M)}(\theta, N_{\varphi_i})+l_{0,h}^{(M)}(\theta, N_{\varphi_i})\leqslant 0\right\}$$

$$=\left\{\frac{1}{r}\left[\sum_{j=0}^{r-1} l_{jm+h}^{(M)}(\theta, N_{\varphi_i})+l_{0,h}^{(M)}(\theta, N_{\varphi_i})\right]\leqslant 0\right\}$$

$$\subset\left\{\frac{1}{r}\left[\sum_{j=0}^{r-1} l_{jm+h}^{(M)}(\theta, N_{\varphi_i})+l_{0,h}^{(M)}(\theta, N_{\varphi_i})\right]-E_\theta l_{0,m}^{(M)}(\theta, N_{\varphi_i})\leqslant -\frac{1}{2}E_\theta l_{0,m}^{(M)}(\theta, N_{\varphi_i})\right\}.$$

再由强收敛的充要条件，得

$$P_\theta\left[\bigcup_{n=N}^{\infty}\left\{\sum_{j=0}^{r-1} l_{jm+h}^{(M)}(\theta, N_{\varphi_i})+l_{0,h}^{(M)}(\theta, N_{\varphi_i})\leqslant 0\right\}\right]\to 0 \quad (N\to\infty), \tag{3.11}$$

式中 $i=1, \cdots, k$.

iv）现在来证明 θ 的 MLE $\hat{\theta}_n$ 有 $P_\theta(\hat{\theta}_n\to\theta)=1$. 也即要证：对任意的 $\varepsilon>0$，$P_\theta\left(\bigcup_{n=N}^{\infty}(\|\hat{\theta}_n-\theta\|\geqslant\varepsilon)\right)\to 0$ $(N\to\infty)$. 但

$$\{\|\hat{\theta}_n-\theta\|\geqslant\varepsilon\}\subseteq\left\{\frac{\prod_{i=1}^{n}p(x_i, \theta)}{\sup_{\varphi\in N_\delta^c}\prod_{i=1}^{n}p(x_i, \varphi)}\leqslant 1\right\}$$

$$\subseteq\left\{\min_{1\leqslant t\leqslant k}\inf_{\varphi\in N_{\varphi_t}}\left[\log\frac{\prod_{i=1}^{h}p(x_i, \theta)}{\prod_{i=1}^{n}p(x_i, \varphi)}+\sum_{j=0}^{r-1}\log\left[\frac{\prod_{jm+h+1}^{(1+j)m+h}p(x_i, \theta)}{\prod_{jm+h+1}^{(1+j)m+h}p(x_i, \varphi)}\right]\leqslant 0\right\}\right.$$

$$\subseteq\left\{\min_{1\leqslant t\leqslant k}\left[l_{0,h}(\theta, \varphi_{N_t})+\sum_{j=0}^{r-1}l_{jm+h,(j+1)m+h}(\theta, N_{\varphi_t})\right]\leqslant 0\right\}$$

$$\subseteq\left\{\bigcup_{t=1}^{k}\left[l_{0,h}(\theta, \varphi_{N_t})+\sum_{j=1}^{r-1}l_{jm+h,(j+1)m+h}(\theta, N_{\varphi_t})\leqslant 0\right]\right\}. \tag{3.12}$$

由(3.11)，(3.12)，立即得到

$$P_\theta\Big(\bigcup_{n=N}^{\infty}\{\|\hat\theta_n-\theta\|\geqslant\varepsilon\}\Big)\leqslant P_\theta\Big(t\bigcup_{t=1}^{k}\bigcup_{n=N}^{\infty}\Big[l_{0,h}^{(M)}(\theta,\ \varphi_{N_t})$$

$$+\sum_{j=0}^{r-1}l_{jm+h}^{(M)}(\theta,\ N_{\varphi_t})\leqslant 0\ \Big]\Big)$$

$$\leqslant\sum_{t=1}^{k}\ P_\theta\Big(\bigcup_{n=N}^{\infty}\Big\{\sum_{j=0}^{r-1}l_{jm+h}^{(M)}(\theta,\ N_{\varphi_t})+l_{0,h}^{(M)}(\theta,\ N_{\varphi_t})\leqslant 0\Big\}\Big)$$

$$\to 0\quad(N\to\infty)\tag{3.13}$$

从而定理证毕.

为了使定理便于验证，下面给出有关条件 3) 的两个充分条件：

i) 对一切 $\theta\in\Theta$，有 $-\infty<E_\theta[\log p(X,\theta)]$，并且存在 $m(\geqslant 1)$，$g(x_1,\cdots,x_m)$，使 $\sup\limits_\theta\Big[\log\prod\limits_{i=1}^{m}p(x_i,\theta)\Big]\leqslant g(x_1,\cdots,x_m)$，$a.s.\ \mu$，且对每一个 $\theta\in\Theta$，$E_\theta g(X_1,\cdots,X_m)<\infty$.

或 ii) 存在 $m(\geqslant 1)$，$g(x_1,\cdots,x_m)$，使

$$\sup_\theta\Big|\log\prod_{i=1}^{m}p(x_i,\theta)\Big|\leqslant g(x_1,\cdots,x_m),\quad a.s.\ \mu,$$

且对一切 $\theta\in\Theta$，$E_\theta g(X_1,\cdots,X_m)<\infty$.

定理 3.2 若定理 3.1 的假设条件成立. 又若进一步有：对任意的 $\varepsilon>0$ 和每一个 $\theta\in\Theta$，在 $N_\theta^c=\{\varphi:\|\theta-\varphi\|\geqslant\varepsilon\}\cap\bar{\Theta}$ 中的每个 φ 皆存在某一邻域 N_φ，正数 $\delta>0$，使得当 $-\delta\leqslant t\leqslant 0$，$m\leqslant s<2m$ 时，有 $E_\theta e^{tl_{0,s}(\theta,N_\varphi)}<\infty$. 那么存在 $c>0$，$a>0$，使得 θ 的 MLE $\hat\theta_n$ 满足下式：

$$P_\theta(\|\hat\theta_n-\theta\|\geqslant\varepsilon)\leqslant ce^{-na},\quad\theta\in\Theta.\tag{3.14}$$

证 令 $n=rm+s$ $(m\leqslant s<2m)$，由定理 3.1 的证明，加上本定理的假设条件，存在 k 个邻域 $N_{\varphi_1},\cdots,N_{\varphi_k}$，使得 N_θ^c 被它们所覆盖，且

$$E_\theta l_{0,m}(\theta,\ N_{\varphi_i})>0,\ E_\theta e^{tl_{0,s}(\theta,N_{\varphi_i})}<\infty$$

$$(-\delta\leqslant t\leqslant 0;\ i=1,\ \cdots,\ k).$$

那么由(3.12)，有

$$P_\theta(\|\hat{\theta}_n-\theta\|\geqslant\varepsilon)\leqslant\sum_{i=1}^{k}P_\theta\Big(\sum_{j=0}^{r-1}l_{jm,(j+1)m}(\theta,\ N_{\varphi_i})$$

$$+l_{rm,rm+s}(\theta,\ N_{\varphi_i})\leqslant 0\Big)$$

$$\leqslant\sum_{i=1}^{k}E_\theta\exp\Big\{t\Big(\sum_{j=0}^{r-1}l_{jm,(j+1)m}(\theta,\ N_{\varphi_i})+l_{rm,rm+s}(\theta,\ N_{\varphi_i})\Big\}$$

$$=\sum_{i=1}^{k}(E_\theta e^{tl_{0,m}(\theta,N_{\varphi_i})})^r E_\theta e^{tl_{0,s}(\theta,N_{\varphi_i})},\tag{3.15}$$

其中 $-\delta\leqslant t\leqslant 0$. 若记随机变量 $l_{0,m}(\theta,\ N_{\varphi_i})$的母函数为

$$M_{m,i}(t)=E_\theta e^{tl_{0,m}(\theta,N_{\varphi_i})},$$

由第五章§1(三)知, 它在定义域内是连续可微函数, 又由假设条件知其定义域包含$[-\delta,\ 0]$, 而 $M'_{m,i}(0-)=E_\theta l_{0,m}(\theta,\ N_{\varphi_i})>0$, 故 $M_{m,i}(t)$在 $t=0$ 的左侧是上升的, 从而存在 $t_i\in[-\delta,\ 0)$, 使

$$M_{m,i}(t_i)<M_{m,i}(0)=1\quad(i=1,\ \cdots,\ k).$$

因此, 若记 $A=\max\limits_{1\leqslant i\leqslant k}M_{m,i}(t_i)$, $B=\max\limits_{1\leqslant i\leqslant k}M_{s,i}(t_i)$, 注意 $0<A<1$, $0<B<\infty$, $r=\dfrac{n-s}{m}\geqslant\dfrac{n-2m}{m}=\dfrac{n}{m}-2$, 则由 (3.15), 有

$$P_\theta(\|\hat{\theta}_n-\theta\|\geqslant\varepsilon)\leqslant KA^rB\leqslant Ce^{-an},\tag{3.16}$$

其中可取 $a=|\log A/m|$, $C=KBe^{-2\log A}$.

注 2 从定理证明过程中可以看到, (3.14) 式中的常数 C 与 a 的选择是依赖于 θ 的. 若加上条件

$$A=\sup_{\substack{\theta\in\Theta\\ \theta+\varphi\in\bar{\Theta}}}\inf_t E_\theta e^{tl_{0,m}(\theta,N_{\varphi_i})}<1;$$

$$B=\sup_{t<0,\theta\in\Theta,\theta+\varphi\in\bar{\Theta}}E_\theta e^{tl_{0,s}(\theta,N_{\varphi_i})}<\infty.$$

则这些常数可以选择得与 θ 无关.

二、位置刻度参数 MLE 的强相合性

定理 3.3 设 $f(x)$为一个 p 维欧氏空间上有界的上半连续的密度函数, $\lim\limits_{\|x\|\to\infty}f(x)=0$, 则

i) 位置参数分布族 $\{f(x-\theta),\ \theta\in R^p\}$ 中 θ 的 MLE $\hat{\theta}_n$ 存在

而且强相合. 若进一步假设存在$\delta>0$, 使得$E_0\{f^{-\delta}(X)\}<\infty$, 则定理 3.2 的结论成立.

ii) 对位置刻度参数分布族$\left\{f\left(\frac{x-\theta}{\sigma}\right)\sigma^{-p},\ \theta\in R^p,\ \sigma>0\right\}$, 如果$|x_1\cdots x_p|^2 f(x_1, \cdots, x_p)$有界, 则$(\theta, \sigma)$的 MLE 存在而且强相合, 并且定理 3.2 成立.

iii) 对刻度参数族$\left\{f\left(\frac{x_1}{\sigma_1}, \cdots, \frac{x_p}{\sigma_p}\right)\frac{1}{\sigma_1}\cdots\frac{1}{\sigma_p},\ \sigma_j>0; j=1, \cdots, p\right\}$, 若存在$\delta>0$, 使得$|x_1\cdots x_p|^{1+\delta}f(x)$有界, 则$\sigma=(\sigma_1, \cdots, \sigma_p)'$的 MLE $\hat{\sigma}_n$存在, 而且强相合, 定理 3.2 的结论也成立.

证明从略, 读者可直接验证定理 3.1 及定理 3.2 的条件. 特别是验证定理 3.1 后指出的有关条件 3) 的两个充分条件. 此证明留给读者作习题.

三、指数族参数的 MLE 的强相合性

定理 3.4 令$X^{(1)}, X^{(2)}, \cdots$为 iid. 随机向量, 服从p维非退化的指数族分布, 具有关于σ有限测度μ的概率密度$p(x, \theta)=\beta(\theta)e^{\theta' T(x)}$, $\theta\in\Omega$, Ω为自然参数区域, 设为开域. 象在 § 2 那样, C为$T(x)$的值域的凸闭包, C^0为C的内部. 如果存在n_0, 使得当$n\geqslant n_0$时, $\bar{T}_n=\frac{1}{n}\sum_{i=1}^{n}T(x^{(i)})$, 对每一个$\theta\in\Omega$, $P_\theta(\bar{T}_n\in(C-C^0))=0$, 则$\theta$、$m(\theta)=E_\theta X$的MLE $\hat{\theta}_n(x^{(1)}, \cdots, x^{(n)})$, $\hat{m}_n(x^{(1)}, \cdots, x^{(n)})$是强相合的, 而且定理 3.2 的结论成立; 又若定理 2.2 的条件成立, 则$m(\theta)$的 MLE 也是强相合的, 此时定理 3.2 的结论也成立.

证 利用定理 2.1, 有关$m(\theta)$的 MLE 为$\bar{T}_n$, 利用 Kolmogorov 强大数定理可知$\bar{T}_n\to m(\theta)\ a.\ s.\ P_\theta$. 由于$m(\theta)$与$\theta$一一对应, 是连续映照. 故$\theta$的$\hat{\theta}_n$也是强相合的. 至于此时定理 3.2 的结论成立, 可直接利用 Chernoff 定理证明. 至于定理的第二部分, 可直接证明. 具体证明步骤, 留给读者作习题.

从定理 3.3 及定理 3.4, 不难验证一些常见的分布族的未知

参数的极大似然估计皆是强相合的，同时估计与真值的欧氏距离大于任意给定的 $\varepsilon>0$ 的概率随抽样个数 n 的增大以指数速度下降为零. 这些常见的分布族是一元和多元正态分布族（估计均值与协方差阵）、Cauchy 分布族、双指数分布族、Γ 分布族、logist 分布族、极值分布族（估计位置，刻度参数），以及二项分布族、Poisson 分布族(估计均值)等等. 这样，是否在所有分布族中未知参数的 MLE 皆有如上的性质呢？并非如此. 试看下例：

例 3.1 设 $Z_i=\begin{bmatrix} X_i \\ Y_i \end{bmatrix}\sim N\left(\begin{pmatrix} \mu_i \\ \mu_i \end{pmatrix}, \sigma^2 I_2\right)$ $(i=1,2,\cdots)$，而且互相独立. 此处 $-\infty<\mu_i<\infty$ $(i=1, 2, \cdots)$. 此时不难求得 σ^2 的无偏估计为

$$\hat{\sigma}_n^2=\frac{1}{2n}\sum_{i=1}^{n}(x_i-y_i)^2,$$

它是强相合的. 而 σ^2 的极大似然估计是

$$\hat{\sigma}_{n,M}^2=\frac{1}{4n}\sum_{i=1}^{n}(x_i-y_i)^2.$$

因为 $P\left[\hat{\sigma}_{n,M}^2\to\frac{\sigma^2}{2}\right]=1$，可见它不强收敛于 σ^2. $\hat{\sigma}_{n,M}^2$ 不具有强相合性的原因，在于无用参数过多. 关于这类 MLE 的讨论，可参阅 Kiefer, Wolfowitz 的工作[11].

例 3.2 设 $X_1, X_2, \cdots$ 为 iid. 随机序列. X_1 具有两点分布

$$P_\theta\{X_1=1\}=\begin{cases}\theta, & \text{若 } \theta \text{ 是有理数;} \\ 1-\theta, & \text{若 } \theta \text{ 是无理数.}\end{cases} \quad (0<\theta<1)$$

θ 的 MLE 为 $\hat{\theta}_n=\frac{1}{n}\sum_{i=1}^{n}x_i$.

$$\hat{\theta}_n\xrightarrow{a.s.P_\theta}\begin{cases}\theta, & \text{当 } \theta \text{ 为有理数;} \\ 1-\theta, & \text{当 } \theta \text{ 为无理数.}\end{cases}$$

这说明 θ 的 MLE $\hat{\theta}_n$ 不是相合估计，其原因在于概率函数对 θ 没有连续性.

§4 极大似然估计的最优渐近正态性

现在我们来研究 MLE 的最优渐近正态性(B. A. N)问题，这个问题一直为一些有名的数理统计学者研究着，例如 Fisher[12]，Dugué[13]，Huber[8]等. 关于这方面的结果，许多数理统计书中都有所叙述. 在这里我们分单参数与多参数两种情况来讨论：在讨论单参数情况时，我们采用了 Danies[14] 的非正规条件，在下面两节中我们进一步讨论 Bahadur、Cramér 渐近有效性时，还要用到它. 在讨论多参数情况时，我们仍然采用一般教科书中所使用的正规条件，其证明较为简单. 如果读者对非正规条件下的结果有兴趣，可以参考[8].

一、单参数情况

设 $X_1, X_2, \cdots$，为 $iid.$ 随机序列，具有分布函数 $F_\theta(x)$，$\theta\in\Theta$，Θ 是实数轴上非退化的开区间，且 $dF_\theta(x)=f(x, \theta)d\mu$，其中 μ 为样本空间 $(\mathscr{X}, \mathscr{B}_x)$ 上 σ-有限测度. $g(\theta)$ 是 Θ 上具有连续导数的实函数，记 $G=g(\Theta)$，为 g 的值域.

$$I(\theta)=E_\theta\left[\frac{\partial}{\partial\theta}\log f(x, \theta)\right]^2$$

为分布族 $\{F_\theta\}$ 的 Fisher 信息函数.

假设 1°) 当 $\theta_1\neq\theta_2(\in\Theta)$ 时，$\mu\{x: f(x_1, \theta_1)\neq f(x, \theta_2)\}>0$. 分布族有共同支撑，不失一般性，可假设对所有的 $x\in\mathscr{X}$，$\theta\in\Theta$，$f(x, \theta)>0$.

2°)*) 记 $Z(x, \theta)=\frac{\partial}{\partial\theta}\log f(x, \theta)$，对每一个 $x\in\mathscr{X}$ 存在，有限，而且是 θ 的连续非增函数，有变号. 对每个 θ，$\mu\{x: Z(x,\theta)\neq 0\}>0$.

3°) 对每个 $\theta_0\in\Theta$，存在 θ_0 的一个邻域 $V_{\theta_0}\subset\Theta$ 及两个函数 $A(x, \theta_0)>0$ 及 $B(x, \theta_0)>0$，使得

*) 此条件可减弱为 $a.s.\mu$ 成立.

$$\left|\frac{f(x,\ \theta)}{f(x,\ \theta_0)}-1\right|\leqslant A(x,\ \theta_0)\,|\theta_0-\theta|\quad (\text{当 }\theta\in V_{\theta_0});\qquad (4.1)$$

$$|Z(x,\ \theta)|\leqslant B(x,\ \theta_0)\qquad (\text{当 }\theta\in V_{\theta_0}).\qquad (4.2)$$

而且 $\quad E_{\theta_0}\{A(X,\ \theta_0)\}<\infty,\ E_{\theta_0}\{B^2(X,\ \theta_0)\}<\infty,$

$$E_{\theta_0}A(X,\ \theta_0)B^2(X,\ \theta_0)<\infty.$$

4°) 令 a、b 表示 Θ 的左右端点，假设 $\lim\limits_{\theta\to a\text{ 或 }b} g(\theta)$ 存在(a、b 可以是 $-\infty$、∞).

定理 4.1 若上述假设条件 1°～3° 成立，则 θ 的 MLE $\hat{\theta}_n(x_1,\ \cdots,\ x_n)$ 是 B. A. N 估计.

在证明此定理之前先证明一条引理. 下面用到的运算记号 P、E 都是指在真参数值 θ_0 下的概率、期望，为行文简便，我们把记号 P_{θ_0}、E_{θ_0} 中的下标 θ_0 略去.

引理 4.1 若上面的假设条件 1°～3° 成立，则有

i) $E(Z(X,\ \theta_0))=0;$ $\qquad (4.3)$

ii) $I(\theta_0)>0$, θ_0 在 Θ 上变动时, $I(\theta_0)$ 连续;

iii) $EZ(X,\ \theta)=-I(\theta_0)(\theta-\theta_0)+o(|\theta-\theta_0|);$ $\qquad (4.4)$

iv) $EZ^2(X,\ \theta)=I(\theta_0)(1+o(1));$ $\qquad (4.5)$

v) $$I(\theta,\ \theta_0)\triangleq E_\theta\log\frac{f(x,\ \theta)}{f(x,\ \theta_0)}=\frac{1}{2}I(\theta_0)(\theta-\theta_0)^2(1+o(1)).\qquad (4.6)$$

其中 iii)～v) 均是指 $V_{\theta_0}\ni\theta\to\theta_0$ 而言.

证 由假设条件 2° 知 $I(\theta_0)>0$; 又由假设条件 3°, $Z^2(x,\theta)\leqslant B^2(x,\ \theta_0)$, 而且 $EB^2(x,\ \theta_0)<\infty$; 再由假设 2°, $Z^2(x,\ \theta)$ 关于 θ 连续. 因此，由控制收敛定理，得

$$\lim_{\theta\to\theta_0}EZ^2(X,\ \theta)=E\lim_{\theta\to\theta_0}Z^2(X,\ \theta)=EZ^2(X,\ \theta_0)=I(\theta_0).$$

从而 iv) 得证. 至于其余结论均是第五章 §1 引理 1.1 的特殊情况，在此不再赘述. 引理证毕.

定理的证明 记 $x=(x_1,\ \cdots,\ x_n)'$, 根据 $\hat{\theta}_n$ 为 θ 的 MLE 的含义，由假设 2°, $Z(x,\ \theta)$ 为 θ 的非增函数，故有

$$\left\{x:\sum_{i=1}^{n} Z(x_i,\ \theta)<0\right\}\subseteq\{x:\hat{\theta}_n<\theta\}\subseteq\left\{x:\sum_{i=1}^{n} Z(x_i,\ \theta)\leqslant 0\right\},\tag{4.7}$$

因而

$$P\left\{\sum_{i=1}^{n} Z(X_i,\ \theta)<0\right\}\leqslant P\{\hat{\theta}_n<\theta\}\leqslant P\left(\sum_{i=1}^{n} Z(X_i,\ \theta)\leqslant 0\right).$$

对任意的 t，选择 $\theta_n\in\Theta$，使得 $\sqrt{nI(\theta_0)}\,(\theta_n-\theta_0)=t$. 因为 $I(\theta_0)>0$, Θ 是开区间，所以当 n 充分大时，θ_n 是存在的，而且当 $n\to\infty$ 时，$\theta_n\to\theta_0$. 由引理 4.1,

$$EZ(X_i,\ \theta_n)=-(\theta_n-\theta_0)I(\theta_0)(1+o(1));$$

$$\begin{aligned}\mathrm{Var}Z(X_i,\ \theta_n)&=EZ^2(X_i,\ \theta_n)-[EZ(X_i,\ \theta_n)]^2\\&=I(\theta_0)(1+o(1)).\end{aligned}$$

故有 $$\frac{E\{Z(X_i,\ \theta_n)\}}{\sqrt{\mathrm{Var}\{Z(X_i,\ \theta_n)\}}}=-(\theta_n-\theta_0)I^{\frac{1}{2}}(\theta_0)(1+o(1));$$

$$\begin{aligned}&P\left(\sum_{i=1}^{n} Z(X_i,\ \theta_n)\leqslant 0\right)\\&=P\left\{\sum_{i=1}^{n}\frac{Z(X_i,\ \theta_n)-EZ(X_i,\ \theta_n)}{\sqrt{n\,\mathrm{Var}\{Z(X_i,\ \theta_n)\}}}\right.\\&\left.\leqslant\sqrt{nI(\theta_0)}(\theta_n-\theta_0)(1+o(1))\right\}.\end{aligned}\tag{4.8}$$

利用中心极限定理及 t 的定义，便有

$$\lim_{n\to\infty} P_{\theta_0}\left(\sum_{i=1}^{n} Z(X_i,\ \theta_n)\leqslant 0\right)=\Phi(t).\tag{4.9}$$

同理可证

$$\lim_{n\to\infty} P_{\theta_0}\left(\sum_{i=1}^{n} Z(X_i,\ \theta_n)<0\right)=\Phi(t),\tag{4.10}$$

此处 $\Phi(t)=\dfrac{1}{\sqrt{2\pi}}\displaystyle\int_{-\infty}^{t} e^{-\frac{u^2}{2}}du.$

利用(4.7)、(4.9)、(4.10)，推得

$$\begin{aligned}\lim_{n\to\infty} P(\hat{\theta}_n<\theta_n)&=\lim_{n\to\infty} P(\sqrt{n}\,(\hat{\theta}_n-\theta_0)<\sqrt{n}\,(\theta_n-\theta_0))\\&=\lim_{n\to\infty} P\left(\sqrt{n}\,(\hat{\theta}_n-\theta_0)<\frac{t}{\sqrt{I(\theta_0)}}\right)=\Phi(t).\end{aligned}\tag{4.11}$$

换言之，$\lim\limits_{n\to\infty} P(\sqrt{n}(\hat{\theta}_n-\theta_0)<t)=\Phi(\sqrt{I(\theta_0)}\,t)$.

因此，$\hat{\theta}_n$ 是 θ 的 BAN 估计. 定理 4.1 证完.

读者不难验证一些常见的分布族，例如正态分布族、二项分布族、Poisson 分布族等皆满足定理的条件. 但是 Cauchy 分布族不满足，因为 $Z(x,\theta)$ 不是 θ 的非增函数.

定理 4.2 在假设 1°～4° 的条件下，$g(\theta)$ 的 MLE $\hat{g}_n(x_1,\cdots,x_n)$ 为 BAN 估计，也即 $\sqrt{n}(\hat{g}_n-g(\theta_0))$ 的极限分布为 $N(0,(g'(\theta_0))^2/I(\theta_0))$.

证 因为 Θ 是开区间，$g'(\theta)$ 连续，$\lim\limits_{\theta\to a \text{ 或 } b} g(\theta)$ 存在，故当 $\hat{\theta}_n\in\Theta$ 时，$\hat{g}_n=g(\hat{\theta}_n)$，

$$\hat{g}_n-g(\theta_0)=(\hat{\theta}_n-\theta_0)(g'(\theta_0)+\eta_n). \tag{4.12}$$

此处 $\eta_n=o_p(1)$ $(n\to\infty)$，由定理 4.1 知，$\sqrt{n}(\hat{g}_n-g(\theta_0))$ 与 $\sqrt{n}(\hat{\theta}_n-\theta_0)g'(\theta_0)$ 具有同一极限分布，并且 $\sqrt{n}\,g'(\theta_0)(\hat{\theta}_n-\theta_0)$ 的极限分布为 $N\left(0,\ \dfrac{(g'(\theta_0))^2}{I(\theta_0)}\right)$，从而定理得证.

二、多参数情况

下面讨论在多参数的情况下 MLE 的 BAN 性质. 我们仍假设 $X_1, X_2, \cdots$，为 $iid.$ 随机序列，具有分布函数 $F_\theta(x)$，$\theta\in\Theta$，而 Θ 是 p 维欧氏空间的开区域. 对每一个 $\theta\in\Theta$，有 $dF_\theta(x)=f(x,\theta)d\mu$，其中 μ 为 $(\mathscr{X},\mathscr{B}_x)$ 上的 σ–有限测度. $g(\theta)$ 是 Θ 上的具有连续偏导数的 k 维向量函数，$G=g(\Theta)$ 为其值域. Fisher 信息阵 $I(\theta)=E_\theta\left(\dfrac{\partial}{\partial\theta}\log f(x,\theta)\right)\left(\dfrac{\partial}{\partial\theta}\log f(x,\theta)\right)'$ 存在且有限.

假设 1°) $f(x,\theta)>0$，对所有 $x\in\mathscr{X}$，$\theta\in\Theta$，且 $\mu\{x: f(x,\theta_1)\neq f(x,\theta_2)\}>0$（当 $\theta_1\neq\theta_2(\in\Theta)$）.

2°) 记 $Z(x,\theta)=\dfrac{\partial\log f(x,\theta)}{\partial\theta}$，为 p 维向量，它存在且有限，并有一阶连续偏导数. $I(\theta)$ 正定.

3°) 对每个 $\theta_0\in\Theta$，存在 θ_0 的邻域 V_{θ_0} 以及 $A(x,\theta_0)>0$,

$B(x, \theta_0)>0$, $H(x, \theta_0)>0$, 使

$$\left|\frac{\partial f(x, \theta)}{\partial \theta}\right| \leqslant A(x, \theta_0); \quad \left|\frac{\partial^2 f(x, \theta)}{\partial \theta_i \partial \theta_j}\right| \leqslant B(x, \theta_0),$$

$$(i, j=1, \cdots, p);$$

$$\left|\left(\frac{\partial Z(x, \theta)}{\partial \theta}-\frac{\partial Z(x, \theta_0)}{\partial \theta}\right)\right| \leqslant H(x, \theta_0)\eta(\theta_0-\theta).$$

且 $\int A(x, \theta_0)d\mu<\infty$, $\int B(x, \theta_0)d\mu<\infty$, $E_{\theta_0}H(X, \theta_0)<\infty$, $\lim\limits_{\theta\to\theta_0}\eta(\theta_0, \theta)=0$.

4°) 当 $\tilde{\theta}$ 是 Θ 的边界点时, $\lim\limits_{\theta\to\tilde{\theta}} g(\theta)$ 存在.

定理 4.3 假设条件 1°～3° 成立, 如果 θ 的估计 $\hat{\theta}_n$ 使得

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z(X_i, \hat{\theta}_n) \xrightarrow{P} 0, \ \hat{\theta}_n \xrightarrow{P} 0 \quad (n\to\infty).$$

那么有 $\mathscr{L}_{\theta_0}(\sqrt{n}(\hat{\theta}_n-\theta_0)) \to N(0, I^{-1}(\theta_0))$, 对每一个 $\theta_0\in\Theta$ 成立, 也即 $\hat{\theta}_n$ 为 θ 的 BAN 估计.

证 由假设条件, θ_0 存在邻域 V_{θ_0}, 使在此邻域内可进行 Taylor 展开

$$\sum_{i=1}^{n} Z(X_i, \hat{\theta}_n)=\sum_{i=1}^{n} Z(X_i, \theta_0)+\sum_{i=1}^{n}\frac{\partial Z(X_i, \theta_0)}{\partial \theta_0}(\hat{\theta}_n-\theta_0)$$
$$+\sum_{i=1}^{n} R(X_i, \hat{\theta}_n, \theta_0)(\hat{\theta}_n-\theta_0). \qquad (4.13)$$

这里 $R(X_i, \hat{\theta}_n, \theta_0)$ 是 $P\times P$ 矩阵, 是 Taylor 展开式的余项部分. 由假设条件 3°, R 的第 (k, l) 元素 $R_{kl}(X_i, \hat{\theta}_n, \theta_0)$ 有 $|R_{kl}(x_i, \hat{\theta}_n, \theta_0)| \leqslant H(x_i, \theta_0)\eta$ $(i=1, \cdots, n; k、l=1, \cdots, p)$, 而且当 $\hat{\theta}_n \xrightarrow{P} \theta_0$ 时, $\eta \xrightarrow{P} 0$. 因为 $EH(X_i, \theta_0)<\infty$, 并注意 $X_1, \cdots, X_n$ 为 iid. 故利用大数法则知

$$\frac{1}{n}\sum_{i=1}^{n} H(X_i, \theta_0)\eta \xrightarrow{P} 0 \quad (n\to\infty).$$

因此有

$$\left|\frac{1}{n}\sum_{i=1}^{n} R(X_i, \hat{\theta}_n, \theta_0)\right| \xrightarrow{P} 0 \quad (n\to\infty). \qquad (4.14)$$

再利用假设条件3°，根据控制收敛定理，易知

$$E_{\theta_0}Z(X,\ \theta_0)=\int\frac{\partial\log f(x,\ \theta)}{\partial\theta}\bigg|_{\theta=\theta_0}f(x,\ \theta_0)d\mu$$

$$=\frac{\partial}{\partial\theta}\int f(x,\ \theta)d\mu=0; \tag{4.15}$$

$$E_{\theta_0}\left(\frac{\partial Z_i(X,\ \theta)}{\partial\theta_j}\bigg|_{\theta=\theta_0}\right)=\int\frac{\partial^2\log f(x,\ \theta)}{\partial\theta_j\partial\theta_i}\bigg|_{\theta=\theta_0}f(x,\ \theta_0)d\mu$$

$$=\int\frac{\partial^2 f(x,\ \theta)}{\partial\theta_j\partial\theta_i}d\mu-\int Z_i(x,\ \theta_0)Z_j(x,\ \theta_0)f(x,\ \theta_0)d\mu$$

$$=0-\int Z_i(x,\ \theta_0)Z_j(x,\ \theta_0)f(x,\ \theta_0)d\mu.$$

此处 $Z(x,\ \theta)=(Z_1(x,\ \theta),\ \cdots,\ Z_p(x,\ \theta))'$，$\theta=(\theta_1,\ \cdots,\ \theta_p)'$. 从而得

$$E_{\theta_0}\left(\frac{\partial Z(X_i,\ \theta)}{\partial\theta}\right)_{\theta=\theta_0}=-I(\theta_0).$$

由此，利用大数法则知，当 $n\to\infty$ 时，有

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial Z(X_i,\ \theta)}{\partial\theta}\right)_{\theta=\theta_0}\xrightarrow{P}-I(\theta_0). \tag{4.16}$$

于是由(4.13)、(4.14)、(4.16)及定理的假设得

$$\sqrt{n}(\hat{\theta}_n-\theta_0)=-I^{-1}(\theta_0)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}Z(X_i,\ \theta_0)$$

$$+\frac{1}{\sqrt{n}}I^{-1}(\theta_0)\sum_{i=1}^{n}Z(X_i,\ \hat{\theta}_n)+o_p(1)$$

$$=-I^{-1}(\theta_0)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}Z(X_i,\ \theta)+o_p(1). \tag{4.17}$$

由(4.17)、$\mathrm{Var}Z(X,\ \theta)=I(\theta)$ 及中心极限定理，得

$$\mathscr{L}_{\theta_0}(\sqrt{n}(\hat{\theta}_n-\theta_0))\to N(0,\ I^{-1}(\theta_0)\quad(\text{当 } n\to\infty). \tag{4.18}$$

$$\lim_{n\to\infty}\mathscr{L}_{\theta_0}\{\sqrt{n}(g(\hat{\theta}_n)-g(\theta_0))\}$$

$$=\lim_{n\to\infty}\mathscr{L}_{\theta_0}\left\{\sqrt{n}\left(\frac{\partial g}{\partial\theta_0}\right)_{\theta=\theta_0}(\hat{\theta}_n-\theta_0)\right\}. \tag{4.19}$$

因此，$g(\theta)$ 的 MLE $\hat{g}_n$ 是 B. A. N. 估计.

例4.1 在例2.1中关于多维正态分布函数 $\Phi(B)$ 的 MLE 就是 B. A. N. 估计.

Wolfowitz 首先在 1963[15] 中,在更广泛的渐近有效性的定义下考虑了 MLE 的渐近有效性. 他把估计类限制为一致相合估计类. 经标准化之后极限分布为 $G(x, \theta)$(不一定是正态分布). 这种渐近有效是基于极限分布覆盖真值的概率大小. Wolfowitz 证明了: 不仅是在渐近正态的一致相合估计类中, 就是在更广泛一致相合估计类中, MLE 也是渐近有效的. 随后 Weiss 和 Wolfowitz 作了推广, 并发展为最大概率估计. 读者可参考文献 [16].

Rao[17]进一步研究了渐近有效估计的第二阶性质. 若 $\{T_n\}$ 是一维 θ 的渐近有效估计,且存在常量 $\alpha(\theta)$、$\beta(\theta)$, 使得

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z(X_i, \theta)-\alpha(\theta)-\sqrt{n}\,\beta(\theta)(T_n-\theta) \xrightarrow{P} 0 \quad (n\to\infty), \tag{4.20}$$

而 Rao 用 $(T_n-\theta)$ 的二阶多项式 $\sqrt{n}\,\alpha(\theta)+n\beta(\theta)(T_n-\theta)+nr(\theta)(T_n-\theta)^2$ 来逼近. $\alpha(\theta)$、$\beta(\theta)$ 象 (4.20) 式那样选择; 而 $r(\theta)$ 则选择为 $\left\{\sum_{i=1}^{n} Z(X_i,\theta)-\sqrt{n}\,\alpha(\theta)-n\beta(\theta)(T_n-\theta)-nr(\theta)(T_n-\theta)^2\right\}$ 有极小渐近方差. 此极小渐近方差称为第二阶渐近有效指标. 一个估计如果达到此指标, 则称为第二阶渐近有效估计. Rao[17] iii) 还比较了多项分布各种渐近有效估计的第二阶效率. 最近日本学者[18]还在他们规定的意义下研究了第三阶渐近有效估计.

§ 5* 极大似然估计的 Bahadur 渐近有效性

在这一节中讨论极大似然估计的 Bahadur 渐近有效性. 在上一章中已经给出了这种渐近有效性的定义及其含义, 它最早是由 Bahadur[9] 所提出的, 后来又有一些作者进行了研究. 例如文献 [5]、[19]、[20] 等. 在这里我们直接讨论多参数情况. 同上章一样, 我们把具有 Bahadur 渐近有效性的估计简记为 BAE 估计.

设 $X_1, X_2, \cdots$ 为从样本空间 $(\mathscr{X}, \mathscr{B}_x)$ 上抽得的 iid. 随机变量序列, 分布函数为 $F_\theta(x)$, $\theta\in\Theta$, Θ 是 p 维空间中的开区域. $F_\theta(x)$ 关于 $(\mathscr{X}, \mathscr{B}_x)$ 上的

σ-有限测度 μ 有密度 $dF_\theta(x)=f(x,\theta)d\mu$. 仍以 $\hat{\theta}_n$、$\hat{g}_n$ 表示 θ、$g(\theta)$ 的基于 $x_1,\cdots,x_n$ 的极大似然估计. 我们研究的问题是: 在什么条件下, 它们具有 BAE 性质. 所谓 $\hat{g}_n$ 具有 BAE 性质是指: 对每个 $\theta_0\in\Theta$ 有

$$\liminf_{\varepsilon\to0}\liminf_{n\to\infty}\frac{1}{n\varepsilon^2}\log P_{\theta_0}(\|g(\theta)-\hat{g}_n\|>\varepsilon)$$

$$=-\frac{1}{2}\frac{1}{\|(D(\theta_0))'I^{-1}(\theta_0)D(\theta_0)\|}. \tag{5.1}$$

此处 $I(\theta_0)$ 是 Fisher 信息阵, $D(\theta_0)=\left(\frac{\partial g(\theta)}{\partial\theta}\right)_{\theta=\theta_0}$, 它为 $p\times k$ 矩阵. 假定 $g(\theta)$ 是具有连续偏导数的 k 维向量函数. 在 $g(\theta)=\theta$ 时, 就得到 $\hat{\theta}_n$ 的 BAE 定义.

在这里我们只讨论这种比较狭义的 BAE 性质. 它多适用于有共同支撑的分布族参数的 $\sqrt{n}^{-1}$ 级相合估计. 对于非共同支撑及其它级相合估计的 Bahadur 渐近有效性, 在[20]中有所讨论, 有兴趣的读者可以参阅.

定理 5.1 假设 § 4 中关于多参数情况的假设 1°～4° 成立. 并进一步假设对每个 $\theta_0\in\Theta$, 存在 $\delta>0$, 使得当 $0<t<\delta$ 时, 有下式成立[注]:

$$E_{\theta_0}e^{t|Z(X,\theta_0)|}<\infty; \tag{5.2}$$

$$E_{\theta_0}e^{t\left|\left(\frac{\partial Z(X,\theta)}{\partial\theta}\right)\right|_{\theta=\theta_0}}<\infty; \tag{5.3}$$

$$E_{\theta_0}e^{tH(X,\theta_0)}<\infty. \tag{5.4}$$

如果 θ 的 MLE $\hat{\theta}_n$ 满足条件: 对任意的 $q>0$, 存在 $a(\theta_0)>0$、$b(\theta_0)>0$, 使得

$$P_{\theta_0}(|\hat{\theta}_n-\theta_0|>q)\leqslant b(\theta_0)e^{-a(\theta_0)n}.$$

则 $g(\theta)$ 的极大似然估计 $\hat{g}_n$ 具有 BAE 性质.

在证明本定理之前, 先证明一条引理:

引理 5.1 设给定 $\varepsilon>0$, 则对每个 $0<\lambda<1$, 存在有限个向量 $l_i\in R^p$, $\|l_i\|=1$ $(i=1,\cdots,M)$, 使得 p 维向量 Y 有如下关系:

$$\{Y:\|Y\|\geqslant\varepsilon\}\subset\bigcup_{i=1}^{M}\{Y:|l_i'Y|\geqslant\lambda\varepsilon\}^{*)}.$$

其中 M 和 $l_1,\cdots,l_M$ 只依赖于 λ, 而与 ε 无关.

证 记 $S_\varepsilon=\{Y:\|Y\|=\varepsilon,\ Y\in R^p\}$, 对每一个 $l\in R^p$, $\|l\|=1$, 记

$$D_l^\varepsilon=\{Y:|l'Y|>\lambda\varepsilon,\ \|Y\|=\varepsilon,\ Y\in R^p\}.$$

D_l^ε 是 $\|Y\|=\varepsilon$ 球面上的开集, S_ε 为有界闭集, 而且显然对每个点 $Y\in S_\varepsilon$, 存在某 l, 例如取 $l=Y/\varepsilon$, 使 $Y\in D_l^\varepsilon$. 于是利用 Heine-Borel 定理, 知必存在有限个 $l_1,\cdots,l_M\in R^p$, 使 $S_\varepsilon\subset\bigcup_{i=1}^{M}D_{l_i}^\varepsilon$. 注意: $l_1,\cdots,l_M$ 的选择虽然依赖于

---

*) 符号 $|\cdot|$ 与第五章相同; $|\cdot|$ 指向量、矩阵中元素的绝对值最大者.

λ 和 ε，但因为对某一 $\varepsilon>0$ 选择的 $l_1, \cdots, l_M$，对其它 $\varepsilon_1>0$ 也适合，故实际上 $l_1, \cdots, l_M$ 仅依赖于 λ. 于是有

$$\{Y: \|Y\| \geqslant \varepsilon\} = \bigcup_{\varepsilon_1 \geqslant \varepsilon} \{Y: \|Y\| = \varepsilon_1\} \subset \bigcup_{\varepsilon_1 \geqslant \varepsilon} \left(\bigcup_{i=1}^{M} D_{l_i}^{\varepsilon_1} \right)$$

$$= \bigcup_{i=1}^{M} D_{l_i}^{\varepsilon} \subset \bigcup_{i=1}^{M} \{Y: |l_i' Y| \geqslant \lambda\varepsilon\}.$$

其中利用了显然的关系式 $D_{l_i}^{\varepsilon_1} \subset D_{l_i}^{\varepsilon}$ （当 $\varepsilon \leqslant \varepsilon_1$ 时）. 引理证毕.

定理的证明 分三步进行：第一步证当 $g(\theta)$ 是 θ 的线性实函数时定理成立；第二步证当 $g(\theta)=\theta$ 时定理成立. 最后证定理成立.

第一步：设 $g(\theta)=d'\theta$，$d\neq 0 (\in R^p)$. 对每个 $\theta_0 \in \Theta$，我们选择适当小的 $q>0$，使下式成立：

$$\{\theta: |\theta-\theta_0|<q\} \subset \Theta, \tag{5.5}$$

$$0=\sum_{i=1}^{n} Z(x_i, \hat{\theta}_n) = \sum_{i=1}^{n} Z(x_i, \theta_0) + \left[\sum_{i=1}^{n} \frac{\partial Z(x_i, \theta_0)}{\partial \theta_0} + \sum_{i=1}^{n} R(x_i, \hat{\theta}_n, \theta_0) \right] (\hat{\theta}_n - \theta_0). \tag{5.6}$$

这样的 q 显然是存在的，因为 Θ 是开区域，(5.5)式不难成立. 又因为 $\hat{\theta}_n$ 是 θ 的 MLE，当 $\hat{\theta}_n \in \Theta$ 时，就有 $\sum\limits_{i=1}^{n} Z(x_i, \hat{\theta}_n)=0$. 再利用 Taylor 展开即得 (5.6) 式. 此处 $R(x_i, \hat{\theta}_n, \theta_0)\cdot(\hat{\theta}_n-\theta_0)$ 是 $Z(x_i, \hat{\theta}_n)$ 展开的余项. 根据 $\dfrac{\partial Z(x, \theta)}{\partial \theta}$ 皆是 θ 的连续函数及假设 3°（见 § 4），存在 η 及 $H(x_i, \theta_0)$，使得

$$|R(x_i, \hat{\theta}_n, \theta_0)| \leqslant H(x_i, \theta_0)\eta(\hat{\theta}_n, \theta_0). \tag{5.7}$$

而且 $E_{\theta_0} H(x_i, \theta_0)<\infty \quad (i=1, \cdots, n)$. $\lim\limits_{\hat{\theta}_n \to \theta_0} \eta(\hat{\theta}_n, \theta_0)=0$.

这时(5.6)式可改写为

$$-\frac{1}{n} \sum_{i=1}^{n} Z(x_i, \theta_0) = \left\{ -I(\theta_0) + \frac{1}{n} \left(\sum_{i=1}^{n} \left(\frac{\partial Z(x_i, \theta_0)}{\partial \theta_0} + I(\theta_0) \right) \right) + \frac{1}{n} \sum_{i=1}^{n} R(x_i, \hat{\theta}_n, \theta_0) \right\} (\hat{\theta}_n - \theta_0) \triangleq \{-I(\theta_0)+W_n\}(\hat{\theta}_n-\theta_0). \tag{5.8}$$

此处 $\quad W_n = \dfrac{1}{n} \left\{ \sum\limits_{i=1}^{n} \left(\dfrac{\partial Z(x_i, \theta_0)}{\partial \theta_0} + I(\theta_0) \right) \right\} + \dfrac{1}{n} \sum\limits_{i=1}^{n} R(x_i, \hat{\theta}_n, \theta_0)$.

由于 $I(\theta_0)>0$，可取 α 相当小，使当 $|W_n|<\alpha$ 时，$I(\theta_0)-W_n$ 非奇异. 从(5.8)式有：当 $|W_n|<\alpha$，$|\hat{\theta}_n-\theta_0|<q$ 时，

$$\hat{\theta}_n - \theta_0 = (I(\theta_0)-W_n)^{-1} \frac{1}{n} \sum_{i=1}^{n} Z(x_i, \theta_0) \tag{5.9}$$

成立. 利用此式，我们来估算 $P_{\theta_0}(|g(\hat{\theta}_n-\theta_0)|>\varepsilon)$.

$$P_{\theta_0}(|g(\hat{\theta}_n-\theta_0)|>\varepsilon)\leqslant P_{\theta_0}(|g(\hat{\theta}_n-\theta_0)|>\varepsilon,\ |W_n|<\alpha,$$
$$|\hat{\theta}_n-\theta_0|<q)+P_{\theta_0}(|W_n|\geqslant\alpha,\ |\hat{\theta}_n-\theta_0|<q)+P_{\theta_0}(|\hat{\theta}_n-\theta_0|\geqslant q). \tag{5.10}$$

现在分别对(5.10)式右端的三项进行估算．根据假设条件，存在 $a(\theta_0)>0$、$b(\theta_0)>0$，使下式成立：

$$P_{\theta_0}(|\hat{\theta}_n-\theta_0|>q)\leqslant b(\theta_0)e^{-a(\theta_0)n}. \tag{5.11}$$

进一步缩小 q，(5.5)与(5.6)仍然成立，并要求当 $|\hat{\theta}_n-\theta_0|\leqslant q$ 时，有

$$\eta(\hat{\theta}_n,\ \theta_0)<\frac{\alpha}{2(E_{\theta_0}H(x,\ \theta_0)+1)}. \tag{5.12}$$

而由 W_n 的定义，有

$$P_{\theta_0}\{|W_n|\geqslant\alpha,\ |\hat{\theta}_n-\theta_0|<q\}\leqslant P_{\theta_0}\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial Z(x_i,\ \theta_0)}{\partial\theta_0}\right)+I(\theta_0)\right|\geqslant\frac{\alpha}{2}\right\}$$
$$+P_{\theta_0}\left\{\left|\frac{1}{n}\sum_{i=1}^{n}R(x_i,\ \hat{\theta}_n,\ \theta_0)\right|\geqslant\frac{\alpha}{2},\ |\hat{\theta}_n-\theta_0|<q\right\}. \tag{5.13}$$

由上节定理4.3的证明过程知

$$E_{\theta_0}\left(\frac{\partial Z(x_i,\ \theta_0)}{\partial\theta_0}\right)=-I(\theta_0).$$

再利用 Chernoff 定理（第五章中的定理1.4）及(5.3)，存在 $0<\rho_{kj}(\alpha)<1$ $(k、j=1,\ \cdots,\ p)$，使得

$$P_{\theta_0}\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial Z(x_i,\ \theta_0)}{\partial\theta_0}\right)+I(\theta_0)\right|\geqslant\frac{\alpha}{2}\right\}$$
$$\leqslant\sum_{k,j=1}^{p}P_{\theta_0}\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial Z_k(x_i,\ \theta)}{\partial\theta_j}\right)_{\theta=\theta_0}+I_{kj}(\theta_0)\right|\geqslant\frac{\alpha}{2}\right\}$$
$$\leqslant 2\sum_{k,j=1}^{p}\rho_{kj}^{n}(\alpha)\leqslant 2p^2\max_{1\leqslant k,j\leqslant p}\rho_{kj}^{n}(\alpha)\triangleq 2p^2\rho_1^n,$$

(5.14)此处 $0<\rho_1<1$．(5.14)

由(5.4)、(5.7)、(5.12)式，用 Chernoff 定理，存在 $0<\rho_2(\alpha)<1$，使

$$P_{\theta_0}\left\{\left|\frac{1}{n}\sum_{i=1}^{n}R(x_i,\ \hat{\theta}_n,\ \theta_0)\right|\geqslant\frac{\alpha}{2},\ |\hat{\theta}_n-\theta_0|<q\right\}$$
$$\leqslant P_{\theta_0}\left\{\frac{1}{n}\sum_{i=1}^{n}H(x_i,\ \theta_0)\eta(\hat{\theta}_n,\ \theta_0)\geqslant\frac{\alpha}{2},\ |\hat{\theta}_n-\theta_0|<q\right\}$$
$$\leqslant P_{\theta_0}\left\{\frac{1}{n}\sum_{i=1}^{n}H(x_i,\ \theta_0)\geqslant E_{\theta_0}H(x,\ \theta_0)+1\right\}\leqslant\rho_2^n(\alpha). \tag{5.15}$$

因此由(5.13)～(5.15)，便有

$$P_{\theta_0}\{|W_n|\geqslant\alpha,\ |\hat{\theta}_n-\theta_0|<q\}\leqslant(2p^2+1)\rho_3^n(\alpha). \tag{5.16}$$

此处，$0<\rho_3(\alpha)=\max\{\rho_1(\alpha),\ \rho_2(\alpha)\}<1$，它与 ε 无关．

现在来估算(5.10)右端的第一项：令 $(I(\theta_0)-W_n)^{-1}=I^{-1}(\theta_0)+\widetilde{W}_n$．因为当 $\alpha\to 0$ 时，$W_n\to 0$，故 $\widetilde{W}_n\to 0$．因而可设

$$|\widetilde{W}_n|<\beta_{\theta_0}(\alpha),$$

且
$$\lim_{\alpha\to 0}\beta_{\theta_0}(\alpha)=0.$$
根据(5.9)式，
$$g(\hat{\theta}_n-\theta_0)=g\left(I^{-1}(\theta_0)\frac{1}{n}\sum_{i=1}^{n}Z(x_i,\ \theta_0)\right)$$
$$+g\left(\widetilde{W}_n\frac{1}{n}\sum_{i=1}^{n}Z(x_i,\ \theta_0)\right). \tag{5.17}$$
当 $|W_n|<\alpha,\ |\hat{\theta}_n-\theta_0|<q$ 时，有
$$\left|g\left(\widetilde{W}_n\frac{1}{n}\sum_{i=1}^{n}Z(x_i,\ \theta_0)\right|\leqslant p^2|d||\widetilde{W}_n|\left|\frac{1}{n}\sum_{i=1}^{n}Z(x_i,\ \theta_0)\right|$$
$$\leqslant p^2|d|\beta_{\theta_0}(\alpha)\left|\frac{1}{n}\sum_{i=1}^{n}Z(x_i,\ \theta_0)\right|.$$
现在选任一个 $\lambda>1$，对任意的 $\varepsilon>0$，由(5.17)式及上式，则有
$$P_{\theta_0}\{|g(\hat{\theta}_n-\theta_0)|>\varepsilon,\ |W_n|<\alpha,\ |\hat{\theta}_n-\theta_0|<q\}$$
$$\leqslant P_{\theta_0}\left\{|g(\hat{\theta}_n-\theta_0)|>\varepsilon,\ \left|g\left(\widetilde{W}_n\frac{1}{n}\sum_{i=1}^{n}Z(X_i,\ \theta_0)\right)\right|<\frac{\lambda-1}{\lambda}\varepsilon\right\}$$
$$+P_{\theta_0}\left\{\left|g\left(\widetilde{W}_n\frac{1}{n}\sum_{i=1}^{n}Z(X_i,\ \theta_0)\right)\right|\right.$$
$$\left.>\frac{\lambda-1}{\lambda}\varepsilon,\ |W_n|<\alpha,\ |\hat{\theta}_n-\theta_0|<q\right\}$$
$$\leqslant P_{\theta_0}\left\{\left|g\left(I^{-1}(\theta_0)\frac{1}{n}\sum_{i=1}^{n}Z(X_i,\ \theta_0)\right)\right|\geqslant\frac{\varepsilon}{\lambda}\right\}$$
$$+P_{\theta_0}\left\{\left|\frac{1}{n}\sum_{i=1}^{n}Z(X_i,\ \theta_0)\right|>\frac{(\lambda-1)\varepsilon}{\lambda|d|p^2\beta_{\theta_0}(\alpha)}\right\}. \tag{5.18}$$
注意 $E_{\theta_0}Z(X_i,\ \theta_0)=0,\ E_{\theta_0}(Z(X_i,\ \theta_0))(Z(X_i,\theta_0))'=I(\theta_0)>0,\ g(\theta)=d'\theta$，有
$$E_{\theta_0}\left\{g(I^{-1}(\theta_0)\frac{1}{n}\sum_{i=1}^{n}Z(X_i,\ \theta_0))\right\}=d'I^{-1}(\theta_0)\sum_{i=1}^{n}E_{\theta_0}Z(X_i,\ \theta_0)=0;$$
$$\mathrm{Var}\,\theta_0\left\{g\left(I^{-1}(\theta_0)\frac{1}{n}\sum_{i=1}^{n}Z(X_i,\ \theta_0)\right)\right\}=\frac{1}{n}d'I^{-1}(\theta_0)d.$$
根据假设条件(5.2)，Chernoff 定理及母函数性质 3)(上一章 § 1)，有
$$P_{\theta_0}\left\{\left|g\left(I^{-1}(\theta_0)\frac{1}{n}\sum_{i=1}^{n}Z(X_i,\ \theta_0)\right)\right|>\frac{\varepsilon}{\lambda}\right\}$$
$$\leqslant\min_t(M_{\theta_0}(t)e^{-\frac{\varepsilon}{\lambda}t})+\min_t M_{\theta_0}(t)e^{\frac{\varepsilon}{\lambda}t}$$
$$\leqslant 2e^{-\frac{n\varepsilon^2/\lambda^2}{2d'I^{-1}(\theta_0)d}}+o\left(\frac{\varepsilon^2}{\lambda^2}\right)\triangleq 2\rho_4(n,\ \varepsilon,\ \lambda). \tag{5.19}$$
其中 $M_{\theta_0}(t)=E_\theta e^{tg\left(I^{-1}(\theta_0)\frac{1}{n}\sum_{i=1}^{n}Z(X_i,\theta_0)\right)}$.

同样，因为 $E_{\theta_0}(Z_k(X_i,\ \theta_0))=0,\ D_{\theta_0}(Z_k(X_i,\ \theta_0))=I_{kk}(\theta_0)$（是 $I(\theta_0)$ 的第 k 行 k 列元素） $(k=1,\ \cdots,\ p;\ i=1,\ \cdots,\ n)$，有

$$
\begin{aligned}
& P_{\theta_0}\left\{\left|\frac{1}{n}\sum_{i=1}^{n} Z(X_i,\ \theta_0)\right| > \frac{(\lambda-1)\varepsilon}{\lambda|d|p^2\beta_{\theta_0}(\alpha)}\right\} \\
& = P_{\theta_0}\left\{\left|\frac{1}{n}\sum_{i=1}^{n} Z(X_i,\ \theta_0)\right| (\lambda|d|p^2\beta_{\theta_0}(\alpha)) > \frac{(\lambda-1)\varepsilon}{\lambda}\right\} \\
& \leqslant \sum_{k=1}^{p} P_{\theta_0}\left\{\frac{1}{n}(\lambda|d|p^2\beta_{\theta_0}(\alpha))\left|\sum_{i=1}^{n} Z_k(X_i,\ \theta_0)\right| > \frac{(\lambda-1)\varepsilon}{\lambda}\right\} \\
& \leqslant 2\sum_{k=1}^{p} e^{-\frac{n(\lambda-1)^2\varepsilon^2}{2I_{kk}(\theta_0)\lambda^2|d|^2p^4\beta_{\theta_0}^2(\alpha)}+o\left(\frac{(\lambda-1)^2\varepsilon^2}{\lambda^2}\right)} \\
& \triangleq 2\sum_{k=1}^{p} \rho_{4+k}(n,\ \varepsilon,\ \alpha).
\end{aligned}
\tag{5.20}
$$

其中又利用了母函数性质:

$$
\begin{aligned}
& \min_{t} E_{\theta_0}\left\{e^{\frac{t}{n}(\lambda|d|p^2\beta_{\theta_0}(\alpha))\sum_{i=1}^{n} Z_k(X_i,\ \theta_0)\pm\frac{t(\lambda-1)\varepsilon t}{\lambda}}\right\} \\
& = e^{-\frac{n(\lambda-1)^2\varepsilon^2}{2I_{kk}(\theta_0)\lambda^2|d|^2p^4\beta_{\theta_0}^2(\alpha)}+o\left(\frac{(\lambda-1)^2\varepsilon^2}{\lambda^2}\right)}.
\end{aligned}
\tag{5.22}
$$

注意(5.19)右边的 $\rho_4(n,\ \varepsilon,\ \lambda)$ 与 α 无关;而(5.20)右边 $\rho_{4+k}(n,\ \varepsilon,\ \alpha)$ 却依赖于 α, 但因为 $\alpha\to 0$ 时, $\beta_{\theta_0}(\alpha)\to 0$. 因此可以进一步缩小 α, 同时再缩小 q, 使(5.12)式成立,并使

$$
\max_{1<k\leqslant n} \frac{-(\lambda-1)^2}{I_{kk}(\theta_0)\lambda^2|d|^2p^4\beta_{\theta_0}^2(\alpha)} < \frac{-1}{\lambda^2 d' I^{-1}(\theta_0)d}. \tag{5.23}
$$

因此,这时有 $\rho_{4+k}(n,\ \varepsilon,\ \alpha) < \rho_4(n,\ \varepsilon,\ \lambda)$ $(k=1,\ \cdots,\ p)$. 再注意,当 $\varepsilon\to 0$ 时, $\rho_4(n,\ \varepsilon,\ \lambda)\to 1$. 故当 ε 适当小时, 有 $\rho_3^n(\alpha) < \rho_4(n,\ \varepsilon,\ \lambda)$, $e^{-a(\theta_0)n} < \rho_4(n,\ \varepsilon,\ \lambda)$. 综合(5.10)、(5.11)、(5.16)、(5.19)及(5.20), 有

$$
\begin{aligned}
& P_{\theta_0}(|g(\hat{\theta}_n-\theta_0)|>\varepsilon) \\
& \leqslant (b(\theta_0)+2p^2+1+2+2p)\rho_4(n,\ \varepsilon,\ \lambda) \\
& = (b(\theta_0)+2p^2+3+2p)e^{-\frac{n\varepsilon^2/\lambda^2}{2d'I^{-1}(\theta_0)d}+o\left(\frac{\varepsilon^2}{\lambda^2}\right)}.
\end{aligned}
\tag{5.24}
$$

因此有

$$
\limsup_{\varepsilon\to 0}\limsup_{n\to\infty}\frac{1}{n\varepsilon^2}\log P_{\theta_0}(|g(\hat{\theta}_n-\theta_0)|>\varepsilon) \leqslant -\frac{1}{2d'I^{-1}(\theta_0)d\lambda^2}. \tag{5.25}
$$

注意: 上式左边与 λ 无关,令 $\lambda\downarrow 1$, 就有

$$
\limsup_{\varepsilon\to 0}\limsup_{n\to\infty}\frac{1}{n\varepsilon^2}\log P_{\theta_0}(|g(\hat{\theta}_n-\theta_0)|>\varepsilon) \leqslant -\frac{1}{2d'I^{-1}(\theta_0)d}. \tag{5.26}
$$

这样就完成了第一步的证明.

第二步: 证明当 $g(\theta)=\theta$ 时定理成立. 利用引理 5.1, 当 $0<\lambda^{-1}<1$ 时, 存在 $l_i\in R^p$, $\|l_i\|=1$ $(i=1,\ \cdots,\ M)$, 使得对任意的 $\varepsilon>0$, 有

$$\{\hat{\theta}_n: \|\hat{\theta}_n-\theta_0\|\geqslant\varepsilon\}\subseteq\bigcup_{i=1}^{M}\{\hat{\theta}_n: |l_i'(\hat{\theta}_n-\theta_0)|\geqslant\lambda^{-1}\varepsilon\},$$

故

$$P_{\theta_0}\{\|\hat{\theta}_n-\theta_0\|\geqslant\varepsilon\}\leqslant P_{\theta_0}\left\{\bigcup_{i=1}^{M}(|l_i'(\hat{\theta}_n-\theta_0)|\geqslant\lambda^{-1}\varepsilon)\right\}$$
$$\leqslant p\max_{1\leqslant i\leqslant M}P_{\theta_0}\{|l_i'(\hat{\theta}_n-\theta_0)|\geqslant\lambda^{-1}\varepsilon\}. \tag{5.27}$$

根据定理第一步证明的结果,有

$$\limsup_{\varepsilon\to0}\limsup_{n\to\infty}\frac{1}{n\varepsilon^2}\log P_{\theta_0}\{\|\hat{\theta}_n-\theta_0\|\geqslant\varepsilon\}$$
$$\leqslant-\frac{1}{2\max\limits_{1\leqslant i\leqslant M}\lambda^2 l_i'I^{-1}(\theta_0)l_i}. \tag{5.28}$$

再根据 $\|I^{-1}(\theta_0)\|$ 的定义,得

$$\max_{1\leqslant i\leqslant M}l_i'I^{-1}(\theta_0)l_i\leqslant\max_{\|l\|=1}l'I^{-1}(\theta_0)l=\|I^{-1}(\theta_0)\|. \tag{5.29}$$

注意到 $l_i(i=1, \cdots, M)$ 与 ε 无关,而(5.28)式的左端与 λ 无关,也与 $l_i(i=1, \cdots, M)$ 无关. 故

$$\limsup_{\varepsilon\to0}\limsup_{n\to\infty}\frac{1}{n\varepsilon^2}\log P_{\theta_0}(\|\hat{\theta}_n-\theta_0\|\geqslant\varepsilon)\leqslant\frac{1}{-2\|I^{-1}(\theta_0)\|}. \tag{5.30}$$

第三步: 先证定理对任意满足假设条件的实函数 $g(\theta)$ 成立. 因为对每个 $\theta_0\in\Theta$, 存在适当小的 $q>0$, 使得 $\{\theta: |\theta-\theta_0|<q\}\subseteq\Theta$,

$$g(\theta)-g(\theta_0)=\left(\frac{\partial g(\theta_0)}{\partial\theta_0}+R(\theta, \theta_0)\right)'(\theta-\theta_0),$$

此处 $\lim\limits_{\theta\to\theta_0}R(\theta, \theta_0)=0$. 以下记 $\dot{g}=\dfrac{\partial g(\theta_0)}{\partial\theta_0}$.

进一步缩小 q, 使对某个 $\lambda>1$, 有

$$\|R(\theta, \theta_0)\|<\beta<\sqrt{\frac{(\lambda-1)^2(\dot{g})'I^{-1}(\theta_0)\dot{g}}{\|I^{-1}(\theta_0)\|}},$$

故

$$P_{\theta_0}\{|g(\hat{\theta}_n)-g(\theta_0)|\geqslant\varepsilon\}$$
$$\leqslant P_{\theta_0}\{|(\dot{g})'(\hat{\theta}_n-\theta_0)+(R(\hat{\theta}_n, \theta_0))'(\hat{\theta}_n-\theta_0)|>\varepsilon, |\hat{\theta}_n-\theta_0|<q\}$$
$$+P_{\theta_0}\{|\hat{\theta}_n-\theta_0|\geqslant q\}$$
$$\leqslant P_{\theta_0}\left\{|(\dot{g})'(\hat{\theta}_n-\theta_0)|>\frac{\varepsilon}{\lambda}, |(R(\hat{\theta}_n, \theta_0))'|<\frac{\lambda-1}{\lambda}\varepsilon\right\}$$
$$+P_{\theta_0}\left\{|(R(\hat{\theta}_n, \theta_0))'(\hat{\theta}_n-\theta_0)|>\frac{\lambda-1}{\lambda}\varepsilon, |\hat{\theta}_n-\theta_0|<q\right\}$$
$$+P_{\theta_0}\{|\hat{\theta}_n-\theta_0|\geqslant q\}$$
$$\leqslant P_{\theta_0}\left\{|(\dot{g})'(\hat{\theta}_n-\theta_0)|\geqslant\frac{\varepsilon}{\lambda}\right\}+P_{\theta_0}\left\{\|\hat{\theta}_n-\theta_0\|\geqslant\frac{\lambda-1}{\lambda}\frac{\varepsilon}{\beta}\right\}$$
$$+P_{\theta_0}\{|\hat{\theta}_n-\theta_0|\geqslant q\}. \tag{5.31}$$

由第一步的(5.24)式及第二步的(5.27)式就有

$$P_{\theta_0}\left\{|(\dot{g})'(\hat{\theta}_n-\theta_0)|\geqslant\frac{\varepsilon}{\lambda}\right\}\leqslant(b(\theta_0)+2p^2+3+2p)e^{\frac{-n\varepsilon^2/\lambda^2}{2(\dot{g})'I^{-1}(\theta)\dot{g}}+o\left(\frac{\varepsilon^2}{\lambda^2}\right)},$$

$$P_{\theta_0}\left\{\|\hat{\theta}_n-\theta_0\|\geqslant\frac{\lambda-1}{\lambda}\frac{\varepsilon}{\beta}\right\}\leqslant pe^{-\frac{n\varepsilon^2(\lambda-1)^2}{2\lambda^2\beta^2\|I^{-1}(\theta_0)\|}+o(\varepsilon^2)}$$

$$\leqslant pe^{-\frac{n\varepsilon^2/\lambda^2}{2(\dot{g})'I^{-1}(\theta_0)\dot{g}}+o(\varepsilon^2)}.$$

再由假设条件知:

$$P_{\theta_0}\{|\hat{\theta}_n-\theta_0|\geqslant q\}\leqslant b(\theta_0)e^{-a(\theta_0)n}$$

$$\leqslant b(\theta_0)e^{-\frac{n\varepsilon^2/\lambda^2}{2(\dot{g})'I^{-1}(\theta_0)\dot{g}}+o(\varepsilon^2)}\quad(\text{当 }\varepsilon\text{ 充分小时}).$$

因此由(5.31)及以上各式,有

$$P_{\theta_0}\{|\hat{g}_n-g(\theta_0)|>\varepsilon\}=P_{\theta_0}\{|g(\hat{\theta}_n)-g(\theta_0)|>\varepsilon\}$$

$$\leqslant(2b(\theta_0)+2p^2+3p+3)e^{-\frac{n\varepsilon^2/\lambda^2}{2(\dot{g})'I^{-1}(\theta)(\dot{g})}+o(\varepsilon^2)},$$

其中 $\hat{g}_n$为 $g(\theta)$的 MLE, 由 MLE 的不变性, 所以有 $\hat{g}_n=g(\hat{\theta}_n)$. 对上式两边取 log, 除以 $n\varepsilon^2$, 然后令 $n\to\infty$, $\varepsilon\to 0$, 再令 $\lambda\downarrow 1$, 从而可得

$$\limsup_{\varepsilon\to 0}\limsup_{n\to\infty}\frac{1}{n\varepsilon^2}\log P_{\theta_0}(|\hat{g}_n-g(\theta_0)|>\varepsilon)$$

$$\leqslant-\frac{1}{2(\dot{g})'I^{-1}(\theta_0)(\dot{g})}\quad(\text{当 }\theta_0\in\Theta\text{ 时}).\tag{5.32}$$

当 $g(\theta)$ 为连续可微的 k 维 ($k\leqslant p$) 向量函数时,则同从第一步到第二步的证明办法完全相似,可以证明,对于 $g(\theta)$的 MLE$\hat{g}_n$, 有

$$\limsup_{\varepsilon\to 0}\limsup_{n\to\infty}\frac{1}{n\varepsilon^2}\log P_{\theta_0}(\|\hat{g}_n-g(\theta_0)\|>\varepsilon)$$

$$\leqslant-\frac{1}{2\|(D(\theta_0))'I^{-1}(\theta_0)D(\theta_0)\|}.\tag{5.33}$$

至此定理证完.

§6 极大似然估计的 Cramér 渐近有效性

自从 Cramér 在他的名著[18] 中提出了参数估计的 (Cramér) 渐近有效性以来, MLE 是否具有 Cramér 渐近有效性的问题, 一般数理统计书中都不大涉及, 似乎有了 BAN 性质, Cramér 渐近有效性质是自然结论,其实这是一种误解.首先我们举出一个常见的例子来说明这个问题. 如果样本空间($\mathscr{X}$, $\mathscr{B}_x$)上具有正态分布

族 $N(\mu, \sigma^2)$，在实际应用中常碰到要估计变异系数 $\frac{\sigma}{\mu}$ $(\mu\neq 0)$. 如果 $X_1, X_2, \cdots$ 是从如上母体中抽得的 $iid.$ 随机序列，那么 $\frac{\sigma}{\mu}$ 的 MLE 为 $\frac{\sqrt{S^2/n}}{\bar{x}}$，此处 $\bar{x}=\frac{1}{n}\sum_1^n x_i$，$S^2=\sum_{i=1}^n(x_i-\bar{x})^2$. 由于 $(\sqrt{S^2/n})/\bar{x}$ 在 $\bar{x}=0$ 时是奇点，显然它的均值及方差都不存在，当然更谈不上有 Cramér 渐近有效性的问题. 但是不难从§4中定理 4.4 推出它具有 BAN 性质. 对于非 MLE Cramér 渐近有效性的讨论还有 Sidney 和 Stone[21] 比较好的结果，这里就不细论及了.

本节仅着重讨论单参数分布族的 MLE Cramér 渐近有效性问题. 设 $X_1, X_2, \cdots$ 为 $iid.$ 随机序列，服从分布 $F_\theta(x)$, $\theta\in\Theta$, Θ 是实轴上的一个开区间，且具有关于 σ-有限测度 μ 的概率密度 $f(x, \theta)=dF_\theta(x)/d\mu$.

引理 6.1 若 $X_1, X_2, \cdots$ 为 $iid.$ 实随机变量序列，$EX_1=0$, $EX_1^2=\sigma^2<\infty$, $E|X_1|^{k_0}<\infty$, $k_0\geqslant 3$. 则对所有 x，有

$$\left|P\left(\frac{\sqrt{n}}{\sigma}\bar{X}<x\right)-\Phi(x)\right|$$
$$\leqslant\frac{C(k)}{(1+|x|)^k}\left(\frac{\rho}{\sqrt{n}}+\frac{E|X_1|^k}{\sigma^k n^{\frac{k-2}{2}}}\right). \qquad (6.1)$$

此处 $\bar{X}=\frac{1}{n}\sum_1^n X_i$, $\rho=\sigma^{-3}E|X_1|^3$, $C(k)$ 是依赖于 $k\leqslant k_0$ 的常数；$\Phi(x)$ 是标准正态分布函数.

由于此引理的证明过于繁琐冗长，因此在此证明从略，读者可参看 Petrov 的著作[22]. 但要指出，此引理的结论是很重要的.

记
$$Z(x, \theta)=\frac{\partial}{\partial\theta}\log f(x, \theta),$$
$$\tilde{Z}(x, \theta)=Z(x, \theta)-E_{\theta_0}Z(x, \theta)$$
$$L(\theta, \theta_0)=E_{\theta_0}\{Z(X, \theta)\}/(D_{\theta_0}\{Z(X, \theta)\})^{\frac{1}{2}},$$
$$\rho_j(\theta, \theta_0)=(E_{\theta_0}|\tilde{Z}(X, \theta)|^j)/(D_{\theta_0}\{Z(X, \theta)\})^{j/2}.$$

定理 6.1 如果本章§4定理 4.1 的假设条件成立，并进一步假设：

(1) 存在某 $k\geqslant 3$ 及 $1>\xi_k>0$, $\eta_k^{(j)}\geqslant 0$, 对所有的 $\theta_0\in\Theta$, 使得

1° $E_{\theta_0}|Z(X,\theta)|^k<\infty$, 对所有的 $\theta\in\Theta$ 成立;

2° 对某一个 $\delta>0$, 存在 $A_\delta>0$, $D_\delta^{(j)}>0$, 使得

$$\inf_{\substack{|\theta-\theta_0|\geqslant\delta\\ \theta\in\Theta}}|\theta-\theta_0|^{-\xi_k}|L(\theta,\theta_0)|\geqslant A_\delta; \tag{6.2}$$

$$\sup_{\substack{|\theta-\theta_0|\geqslant\delta\\ \theta\in\Theta}}\frac{\rho_j(\theta,\theta_0)}{|L(\theta,\theta_0)|^{\eta_k^{(j)}}}\leqslant D_\delta^{(j)}<\infty\quad (j=3,k). \tag{6.3}$$

3° $$\min_{j=3,k}\xi_k(k-\eta_k^{(j)})-2>0. \tag{6.4}$$

或者 (2) 若有 $t_0>0$ 及 $\alpha>0$, 对所有的 $\theta_0\in\Theta$, 存在适当的 $\delta>0$ 及 $B_\delta>0$, 使得当 $\theta_0\leqslant\theta$, $0\leqslant t\leqslant t_0$ 或 $\theta\leqslant\theta_0$, $-t_0<t\leqslant 0$ 时,

$$M_\theta(t)\triangleq E_{\theta_0}e^{tZ(x,\theta)}<\infty. \tag{6.5}$$

而且存在适当大的 m, 使得 $\dfrac{(m\delta)^\alpha}{B_\delta}>1$, 且

$$\sup_{|\theta-\theta_0|\geqslant\delta_m}|\theta-\theta_0|^\alpha\inf_t M_\theta(t)\leqslant B_\delta<\infty. \tag{6.6}$$

在 (1)、(2) 两组条件之一成立时, 则 θ 的 MLE 为 CAE 估计.

证 首先对(1)组条件来证: 令

$$\begin{aligned} H_{n,k}(\theta)&\triangleq\frac{C(k)}{(1+\sqrt{n}|L(\theta,\theta_0)|)^k}\\ &\quad\times\left(\frac{1}{\sqrt{n}}\rho_3(\theta,\theta_0)+n^{-\frac{k-2}{2}}\rho_k(\theta,\theta_0)\right);\\ G_{n,\theta_0}(x)&\triangleq P_{\theta_0}(\sqrt{n}|\hat\theta_n-\theta_0|>x);\\ x&=\sqrt{n}|\theta-\theta_0|. \end{aligned} \tag{6.7}$$

因为 $\hat\theta_n$ 是 θ 的 MLE 以及 $Z(x,\theta)$ 是 θ 的单调非增函数, 故

$$\begin{aligned}\left\{(x_1,\cdots,x_n)':\sum_{i=1}^n Z(x_i,\theta)<0\right\}&\subseteq\{(x_1,\cdots,x_n)':\hat\theta_n<\theta\}\\ &\subseteq\left\{(x_1,\cdots,x_n)':\sum_{i=1}^n Z(x_i,\theta)\leqslant 0\right\}\end{aligned} \tag{6.8}$$

根据引理 6.1, 当 $\theta_0\pm|\theta-\theta_0|\in\Theta$ 时,

$$
\begin{aligned}
&|G_{n,\theta_0}(x)-\Phi(\sqrt{n}\,L(\theta_0+|\theta-\theta_0|,\ \theta_0))\\
&\quad-\Phi(-\sqrt{n}\,L(\theta_0-|\theta-\theta_0|,\ \theta_0))|\\
&\quad\leqslant H_{n,k}(\theta_0+|\theta-\theta_0|)+H_{n,k}(\theta_0-|\theta-\theta_0|).
\end{aligned}
$$

由于 $Z(x,\theta)$ 是 θ 的非增函数，由引理 4.1，$E_{\theta_0}Z(X,\theta_0)=0$，$I(\theta_0)>0$，故

$$
E_{\theta_0}\{Z(X,\theta)\}\begin{cases}<0, & 当\ \theta>\theta_0;\\ >0, & 当\ \theta<\theta_0.\end{cases} \tag{6.9}
$$

当 $\theta>\theta_0$ 时，$L(\theta,\theta_0)<0$；当 $\theta<\theta_0$ 时，$L(\theta,\theta_0)>0$. (6.10)

下面分两种情况来分析：先考虑 $|\hat{\theta}_n-\theta_0|\geqslant\delta$ 的情况：由 (1) 组的假设 2'，当 $|\theta-\theta_0|\geqslant\delta$ 时，就有

$$
\left.\begin{aligned}
L(\theta,\ \theta_0)&\geqslant A_\delta|\theta-\theta_0|^{\xi_k}, \quad 当\ \theta<\theta_0;\\
-L(\theta,\ \theta_0)&\geqslant A_\delta|\theta-\theta_0|^{\xi_k}, \quad 当\ \theta>\theta_0.
\end{aligned}\right\} \tag{6.11}
$$

$$
\begin{aligned}
H_{n,k}(\theta_0\pm|\theta-\theta_0|)&\leqslant C(k)\Big\{n^{-\frac{1}{2}}D_\delta^{(3)}|L(\theta_0\pm|\theta-\theta_0|,\ \theta_0)|^{\eta_k^{(3)}}\\
&\qquad+n^{-\frac{k-2}{2}}D_\delta^{(k)}|L(\theta_0\pm|\theta-\theta_0|,\ \theta_0)|^{\eta_k^{(k)}}\Big\}\Big/\\
&\qquad n^{\frac{k}{2}}|L(\theta_0\pm|\theta-\theta_0|,\ \theta_0)|^k\\
&\leqslant C(k)\left[\frac{D_\delta^{(3)}}{n^{\frac{k+1}{2}}|L(\theta_0\pm|\theta-\theta_0|,\ \theta_0)|^{k-\eta_k^{(3)}}}\right.\\
&\qquad\left.+\frac{D_\delta^{(k)}}{n^{k-1}|L(\theta_0\pm|\theta-\theta_0|,\ \theta_0)|^{k-\eta_k^{(k)}}}\right]\\
&\leqslant\left(\frac{C(k)D_\delta^{(3)}}{A_\delta^{k-\eta_k^{(3)}}}\right)\frac{1}{n^{\frac{k+1}{2}-\frac{1}{2}\xi_k\left(k-\eta_k^{(3)}\right)}x^{\xi_k\left(k-\eta_k^{(3)}\right)}}\\
&\qquad+\left(\frac{C(k)D_\delta^{(k)}}{A_\delta^{k-\eta_k^{(k)}}}\right)\frac{1}{n^{k-1-\frac{1}{2}\xi_k\left(k-\eta_k^{(k)}\right)}x^{\xi_k\left(k-\eta_k^{(k)}\right)}}.
\end{aligned} \tag{6.12}
$$

因此

$$
\begin{aligned}
&\int_{|\hat{\theta}_n-\theta_0|>\delta}xG_{n,\theta_0}(x)\,dx=\int_{\delta\sqrt{n}}^{\infty}xG_{n,\theta_0}(x)\,dx\\
&\quad\leqslant\int_{\delta\sqrt{n}}^{\infty}2x\Phi(-A\delta n^{\frac{1}{2}-\frac{\xi_k}{2}}x^{\xi_k})\,dx\\
&\qquad+\left(\frac{C(k)D_\delta^{(3)}}{A_\delta^{k-\eta_k^{(3)}}}\right)\frac{1}{n^{\frac{k-1}{2}}}\int_{\delta\sqrt{n}}^{\infty}\frac{1}{n^{1-\frac{1}{2}\xi_k\left(k-\eta_k^{(3)}\right)}}\cdot\frac{dx}{x^{\xi_k\left(k-\eta_r^{(3)}\right)-1}}
\end{aligned}
$$

$$+\left(\frac{C(k)D_\delta^{(k)}}{A_\delta^{k-\eta_k^{(k)}}}\right)\frac{1}{n^{k-2}}\int_{\delta\sqrt{n}}^{\infty}\frac{1}{n^{1-\frac{1}{2}\xi_k\left(k-\eta_k^{(k)}\right)}}\cdot\frac{dx}{x^{\xi_k\left(k-\eta_k^{(k)}\right)-1}}$$

$$=o(n^{-\frac{k-1}{2}})+\left(\frac{C(k)D_\delta^{(3)}}{A_\delta^{k-\eta_k^{(3)}}}\right)\frac{1}{n^{\frac{k-1}{2}}}\int_\delta^\infty z^{-\xi_k\left(k-\eta_k^{(3)}\right)+1}\,dz$$

$$+\left(\frac{C(k)D_\delta^{(k)}}{A_\delta^{k-\eta_k^{(k)}}}\right)n^{-k+2}\int_\delta^\infty z^{-\xi_k\left(k-\eta_k^{(k)}\right)+1}\,dz$$

$$=o(n^{-\frac{k-1}{2}}). \tag{6.13}$$

现在来分析 $|\hat{\theta}_n-\theta_0|\leqslant\delta$ 的情况：由引理 4.1,

$$L(\theta,\ \theta_0)=-(\theta-\theta_0)I^{\frac{1}{2}}(\theta_0)(1+o(1))$$
$$\triangleq-(\theta-\theta_0)b(1+o(1)).$$

故
$$\sqrt{n}\,L(\theta,\ \theta_0)=\pm bx(1+o(1)).$$

这里 $o(1)$ 是指 $\frac{x}{\sqrt{n}}\to 0$ 而言. 因此

$$\frac{bx}{2}<\sqrt{n}\,|L(\theta,\ \theta_0)|<\frac{3b}{2}x,$$

从而

$$\limsup_{n\to\infty}\int_0^{\delta\sqrt{n}}xG_{n,\theta_0}(x)\,dx$$

$$\leqslant\limsup_{n\to\infty}\Bigg[\int_0^L 2x\Phi(-bx(1+o(1)))\,dx$$

$$+\max_{\substack{|\theta-\theta_0|<\delta\\ \theta\in\Theta}}\frac{4C(3)}{\sqrt{n}}\int_0^\infty\frac{x\,dx}{\left(1+\frac{bx}{2}\right)^3}+2\int_L^\infty x\Phi\left(-\frac{b}{2}x\right)dx\Bigg]$$

$$=2\int_0^L x\Phi(-bx)\,dx+2\int_L^\infty x\Phi\left(-\frac{b}{2}x\right)dx. \tag{6.14}$$

上面不等式是利用 $k=3$ 时引理 6.1 的结论. 再令 $L\to\infty$, 即得

$$\limsup_{n\to\infty}\int_0^{\delta\sqrt{n}}xG_{n,\theta_0}(x)\,dx\leqslant 2\int_0^\infty x\Phi(-bx)\,dx=\frac{1}{2b^2}$$

$$=\frac{1}{2I(\theta_0)}. \tag{6.15}$$

用证明(6.15)式的方法同样可证

$$\liminf_{n\to\infty}\int_0^{\delta\sqrt{n}} xG_{n,\theta_0}(x)\,dx \geqslant 2\int_0^L x\Phi(-bx)\,dx$$

$$\xrightarrow[L\to\infty]{} 2\int_0^\infty x\Phi(-bx)\,dx = \frac{1}{2b^2} = \frac{1}{2I(\theta_0)}, \tag{6.16}$$

也即

$$\liminf_{n\to\infty}\int_0^{\delta\sqrt{n}} xG_{n,\theta_0}(x)\,dx \geqslant \frac{1}{2I(\theta_0)}. \tag{6.17}$$

合(6.15)及(6.17)，得

$$\lim_{n\to\infty}\int_0^{\delta\sqrt{n}} xG_{n,\theta_0}(x)\,dx = \frac{1}{2I(\theta_0)}. \tag{6.18}$$

用分部积分法，我们知道

$$E_{\theta_0} n(\hat{\theta}_n-\theta_0)^2 = 2\int_0^\infty xG_{n,\theta_0}(x)\,dx \to I^{-1}(\theta_0). \tag{6.19}$$

至此证明了在(1)组条件成立时，定理为真.

下面证明在(2)组条件下，定理为真：由于在上面证明

$$\lim_{n\to\infty}\int_0^{\delta\sqrt{n}} xG_{n,\theta_0}(x)\,dx = \frac{1}{2I(\theta_0)}$$

时，只利用到§4的定理4.1的条件，而未涉及(1)组条件，因此结论对(2)组条件仍然成立. 余下只要证明：在(2)组条件下，也有 $\lim\limits_{n\to\infty}\int_{\delta\sqrt{n}}^\infty xG_{n,\theta_0}(x)\,dx = 0$ 即可.

由 Chernoff 定理及(2)组的假设，我们仍记

$$x = \sqrt{n}\,|\theta-\theta_0|.$$

当 $|\theta-\theta_0| \geqslant m\delta$ 时，

$$\begin{aligned}
G_{n,\theta_0}(x) &= P_{\theta_0}(\sqrt{n}\,|\hat{\theta}_n-\theta_0| > x) \\
&\leqslant P_{\theta_0}\Big(\sum_{i=1}^n Z(X_i,\ \theta_0-|\theta-\theta_0|) \leqslant 0\Big) \\
&\quad + P_{\theta_0}\Big(\sum_{i=1}^n Z(X_i,\ \theta_0+|\theta-\theta_0|) \geqslant 0\Big) \\
&\leqslant (\inf_t M_{\theta_0-|\theta-\theta_0|}(t))^n + (\inf_t M_{\theta_0+|\theta-\theta_0|}(t))^n \\
&\leqslant 2(B_\delta|\theta-\theta_0|^{-\alpha})^n.
\end{aligned} \tag{6.20}$$

这里 $M_\theta(t) = E_{\theta_0} e^{tZ(x,\theta)}$. 因为由(2)组条件，$I(\theta)>0$，$Z(x,\ \theta)$ 对

每个 x 是 θ 的非增函数，从而可知，$M_\theta(t)$ 是 t 的严凸函数，当 $t>0$ 时，它是 θ 的降函数；当 $t<0$ 时，它是 θ 的升函数，$M_\theta(0)=1$，$M'_\theta(0)=E_{\theta_0}Z(X, \theta)\neq 0$，故存在 t_0，使得 $M_\theta(t_0)<M_\theta(0)=1$，而且

$$\text{当 } |\theta-\theta_0|\geqslant\delta \text{ 时，}\quad \inf_t M_\theta(t)\leqslant\inf_t M_{\theta_0\pm\delta}(t)<1. \tag{6.21}$$

因此

$$\int_{\delta\sqrt{n}}^{\infty} xG_{n,\theta_0}(x)dx\leqslant 2\int_{\delta\sqrt{n}}^{m\delta\sqrt{n}} x[\inf_t M_{\theta_0\pm\delta}(t)]^n dx$$

$$+2\int_{\delta m\sqrt{n}}^{\infty} x(n^{\frac{\alpha}{2}}B_\delta x^{-\alpha})^n dx$$

$$\leqslant m^2\delta^2 n(\inf M_{\theta_0\pm\delta}(t))^n+2\int_{m\delta/B_\delta^{\frac{1}{\alpha}}}^{\infty} B_\delta^{\frac{2}{\alpha}} x^{-n\alpha+1}dx$$

$$\leqslant m^2\delta^2 n\ (\inf M_{\theta_0\pm\delta}(t))^n+2B_\delta^{\frac{2}{\alpha}}\left(\frac{m\delta}{B_\delta^{\frac{1}{\alpha}}}\right)^{-n\alpha+2}\frac{n}{n\alpha-2}$$

$$\to 0 \quad (n\to\infty).$$

至此定理证毕.

系 6.1 如果 § 4 的定理 4.1 的假设成立，则有

$$\lim_{n\to\infty} F_{\theta_0}\{n(\hat{\theta}_n-\theta_0)^2 I_{[|\hat{\theta}_n-\theta_0|<\delta]}(X_1, \cdots, X_n)\}$$

$$=\lim_{n\to\infty} n\int_{|x|<\delta\sqrt{n}} x^2 dP_{\theta_0}(\sqrt{n}(\hat{\theta}_n-\theta_0)<x)=I^{-1}(\theta_0). \tag{6.22}$$

而满足此关系的 θ 的估计 $\hat{\theta}_n$ 称为弱 Cramér 有效估计.

系 6.2 当 X 服从单参数指数族分布

$$f(x, \theta)d\mu=\beta(\theta)e^{\theta x}d\mu, \quad \theta\in\boldsymbol{\Theta}=\Theta^0 \quad (\text{开区间})$$

时，记 $m(\theta)=E_\theta X$，如果存在 $\delta>0$，及相应的 $\alpha>0$，使

$$\sup_{\substack{|\theta-\theta_0|\geqslant\delta\\ \theta\in\Theta}} |\theta-\theta_0|^{-\alpha}e^{-m(\theta)(\theta-\theta_0)}\beta(\theta_0)\beta^{-1}(\theta)\leqslant B_\delta<\infty, \tag{6.23}$$

则 θ 的 MLE 具有 Cramér 渐近有效性.

证 容易验证对指数族分布在 $\theta\in\Theta^0$ 时，定理 4.1 的假设条件得到满足，这在上节已作了分析. 下面主要验证定理的(2)组条件成立：

因为$Z(x, \theta)=x-m(\theta)$,

$$M_\theta(t)=E_{\theta_0}e^{tZ(X,\theta)}=E_{\theta_0}e^{t(X-m(\theta))}=e^{-m(\theta)t}\beta(\theta_0)\beta^{-1}(\theta_0+t),$$

由于θ_0是Θ的内点，故存在$t_0>0$, 使得$(\theta_0-t_0, \theta_0+t_0)\subset\Theta$, 因而当$t\in(\theta_0-t_0, \theta_0+t_0)$时，有$M_\theta(t)<\infty$, 此结果对一切$\theta_0\in\Theta$成立. 故

$$\inf_t M_\theta(t)=e^{-m(\theta)(\theta-\theta_0)}\beta(\theta_0)\beta^{-1}(\theta).$$

因此定理的(2)组条件成立. 定理中的$m\delta$相当此处的δ. 注意: 假设条件中对某$\delta>0$及B_δ, (6.23)式成立, 则对$\delta_1>\delta$及B_δ, (6.23)式更成立, 因此总可假定$\frac{\delta^\alpha}{B_\delta}>1$.

定理 6.2 如果定理4.1的假设成立, $\hat\theta_n$是θ的MLE, 则对$0<\eta<1$, 令

$$\tilde\theta_n=\begin{cases}\hat\theta_n, & \text{当 } |\hat\theta_n|\leqslant n^{\frac{1-\eta}{2}};\\ n^{\frac{1-\eta}{2}}\operatorname{Sgn}\hat\theta_n, & \text{其它}.\end{cases} \tag{6.24}$$

则$\hat\theta_n$是Cramér渐近有效估计.

证 由上面的系6.1知, $\hat\theta_n$是弱Cramér渐近有效估计, 也即对任意的$\theta_0\in\Theta$, 有

$$\int_0^{\delta\sqrt{n}} x^2 dP_{\theta_0}(\sqrt{n}\,|\hat\theta_n-\theta_0|<x)\to I^{-1}(\theta_0)\quad(n\to\infty). \tag{6.25}$$

因此当n充分大时, 下式成立:

$$\{|\hat\theta_n|\leqslant n^{\frac{1-\eta}{2}}\}\supset\{|\hat\theta_n-\theta_0|\leqslant\delta\};$$

$$\begin{aligned}E_{\theta_0}[n(\tilde\theta_n-\theta_0)^2]&=E_{\theta_0}\{n(\hat\theta_n-\theta_0)^2I_{[|\hat\theta_n-\theta_0|<\delta]}\}\\&\quad+E_{\theta_0}\{n(\tilde\theta_n-\theta_0)^2I_{[|\hat\theta_n-\theta_0|>\delta]}\\&=\int_0^{\delta\sqrt{n}}x^2dP_{\theta_0}(\sqrt{n}\,|\hat\theta_n-\theta_0|<x)+R_n,\end{aligned} \tag{6.26}$$

其中

$$\begin{aligned}|R_n|&\leqslant n(n^{\frac{1-\eta}{2}}+|\theta_0|)^2P_{\theta_0}(|\hat\theta_n-\theta_0|>\delta)\\&\leqslant n(n^{\frac{1-\eta}{2}}+|\theta_0|)^2[\Phi(\sqrt{n}L(\theta_0+\delta, \theta_0))\\&\quad+\Phi(-\sqrt{n}L(\theta_0-\delta, \theta_0))\\&\quad+H_{n,3}(\theta_0+\delta)+H_{n,3}(\theta_0-\delta).\end{aligned} \tag{6.27}$$

根据定理 6.1 及本定理的假设,

$$E_{\theta_0}Z(X,\ \theta_0+\delta)<0;\quad E_{\theta_0}Z(X,\ \theta_0-\delta)>0. \tag{6.28}$$

从而 $L(\theta_0+\delta,\ \theta_0)<0$; $L(\theta_0-\delta,\ \theta_0)>0$. 又有

$$H_{n,3}(\theta_0+\delta)=\frac{2C(3)\rho_3(\theta_0+\delta)}{\sqrt{n}\,(1+\sqrt{n}\,|L(\theta_0+\delta,\ \theta_0)|)^3}=O(n^{-2}); \tag{6.29}$$

$$H_{n,3}(\theta_0-\delta)=\frac{2C(3)\rho_3(\theta_0-\delta)}{\sqrt{n}\,(1+\sqrt{n}\,|L(\theta_0-\delta,\ \theta_0)|)^3}=O(n^{-2}). \tag{6.30}$$

众所周知, 对任意的 $l>0$, 当 $x\to\infty$ 时, $x^l\Phi(-x)\to0$. 因此由(6.28)及此结论,有

$$\left.\begin{aligned}&n^l\Phi(\sqrt{n}\,L(\theta_0+\delta,\ \theta_0))\to0\quad(\text{当 } n\to\infty);\\&n^l\Phi(-\sqrt{n}\,L(\theta_0-\delta,\ \theta_0))\to0\quad(\text{当 } n\to\infty).\end{aligned}\right\} \tag{6.31}$$

再根据以上的(6.27)～(6.31)及 $0<\eta<1$, 可立即推得

$$R_n\to0\quad(n\to\infty). \tag{6.32}$$

再由(6.25)、(6.26)及(6.32), 就证得

$$nE_{\theta_0}(\tilde{\theta}_n-\theta_0)^2\to I^{-1}(\theta_0)$$

对任意 $\theta_0\in\Theta$ 成立.

此定理可使很大一部分非 Cramér 渐近有效的极大似然估计能改进为 Cramér 渐近有效估计. 甚至 θ 的 MLE 不存在,当由于部分样本点所引起时,这时在大样本理论中往往不起主要作用. 利用上述方法,常可构造出 Cramér 渐近有效估计. 例如 Poisson 分布, 我们要估计 $\theta=\log\lambda$. 在本章的 §2 中已经指出了 θ 的 MLE 不存在. 因为 $\overline{X}=0$ 时, θ 在负无穷大处才使似然函数达到极大. 但是令

$$\tilde{\theta}_n(x)=\begin{cases}\log\bar{x}, & \text{当 } \bar{x}>0;\\ -\log n, & \text{当 } \bar{x}=0.\end{cases}$$

不难证明此估计 $\tilde{\theta}_n(x)$ 是 Cramér 渐近有效估计.

至于多参数的情况,结果少见,这里就不讨论了. 有关多参数指数族的分布、均值及其有关函数估计的 Cramér 渐近有效性, 可

以用上一章有关矩估计渐近性质的讨论来加以证明. 具体证明留给读者作习题.

关于其它有关极大似然估计的讨论, 还有几个方面是值得注意的, 例如群观察资料极大似然估计及其渐近性质的讨论[25]. 在截尾样本情形下的极大似然估计的讨论, 可参考文献[26].

问题与习题

1. 令 $y_{ij}=\xi_i+\eta_j$, $\xi_i\sim N(\theta,\ \sigma_1^2)$, $\eta_j\sim N(0,\ \sigma_2^2)$ $(i=1,\ 2,\ \cdots,\ n;\ j=1,\ 2,\ \cdots,\ m)$. 且 $\xi_1,\ \cdots,\ \xi_n;\eta_1,\ \cdots,\ \eta_m$ 彼此独立. 试求 σ_1^2、σ_2^2 的 MLE. 并证明其均方差比其 UMVUE 一致地小.

在以下题目中皆设 n 是抽样大小, $x_1,\ x_2,\ \cdots,\ x_n$ 是由总体独立抽样的结果. 在题目中不再说明.

2. 试求总体服从极值分布(其密度函数当 $x>\theta$ 时, 为

$$f(x,\ \theta)=e^{-(x-\theta)-e^{-(x-\theta)}};$$

当 $x\leqslant\theta$ 时, $f(x,\ \theta)=0$)θ 的 MLE.

3. 若总体服从 $(\theta,\ \theta+|\theta|)$ $(\theta\in\Theta)$ 的均匀分布, 试求在 $\Theta=(-\infty,\ 0)$、$(0,\ \infty)$ 下 θ 的 MLE 与其 MLE 的方差. 并证明

$$p\,\frac{n+1}{n+2}\min_{1\leqslant i\leqslant n}x_i+(1-p)\max_{1\leqslant i\leqslant n}x_i$$

是 θ 的无偏估计, 并求此估计的方差.

4. 若总体服从指数族分布, 其密度函数当 $x\geqslant1$, $\theta\geqslant0$ 时, 为 $f(x,\ \theta)=\beta(\theta)x^{-4}e^{-\theta x}$; 在其他情况, $f(x,\ \theta)=0$. 试问 θ 的 MLE 满足似然方程 $\beta'(\theta)/\beta(\theta)=\frac{1}{n}\sum\limits_{i=1}^{n}x_i$ 吗?

5. 若总体服从对数正态分布 $(\log X\sim N(\theta,\ \sigma^2))$, 试求总体的均值、方差和其 MLE, 并算出 MLE 的协方差阵与渐近展式.

6. 若总体服从 Poisson 分布 $\mathscr{P}(\lambda)$, 求证 $e^{-\lambda}$ 的 UMVUE 为 $\left(1-\frac{1}{n}\right)^{\sum_1^n x_i}$, MLE 为 $e^{-\frac{1}{n}\sum_1^n x_i}$, 皆有同一极限分布 $N(0,\ \lambda e^{-2\lambda})$.

7*. 若总体具有密度函数 $f(x,\ \theta)=\beta(\theta)e^{\theta' x}r(x)$, 其中 $x\in R^2$, $\Theta=R^2$, $r(x)=1$(当 $\|x\|^2=1$ 时); 在其他情况, $f(x,\ \theta)=0$. 试证明当 $n=1$ 时, θ 的 MLE 不存在, 而 $n\geqslant2$ 时存在.

8. 若总体具有密度函数当 $x>\theta$ 时,

$$f(x;\alpha,\ \beta,\ \theta)=\frac{1}{\beta}(x-\theta)^{\alpha-1}e^{-\frac{(x-\theta)^{\alpha}}{\beta}};$$

在其他情况，$f(x;\alpha,\ \beta,\ \theta)=0$. 其中 $\alpha>0,\ \beta>0,\ \theta\in R^1$. 在已知 $\alpha=\alpha_0$ 时，求$(\theta,\ \beta)$的 MLE. 并证明 θ、α、β 的 MLE 是强相合的.

9. 若总体具有密度函数 $f(x;a,b,\sigma)=\frac{1}{b-a}\left(\Phi\left(\frac{x-a}{\sigma}\right)-\Phi\left(\frac{x-b}{\sigma}\right)\right)$，其中 $b>a,\ \sigma>0$. 求证 a、b、σ 的 MLE 是强相合的，而且是 BAN 估计.

[提示：不妨扩大 $f(x;a,\ b,\ \sigma)$到 a、$b\in R^1$ 情况下讨论.]

10. 若总体服从 Cauchy 分布，其密度为 $\frac{1}{\pi}\cdot\frac{1}{1+(x-\theta)^2}$，令 $x^*_{[\frac{n}{2}]}$ 表示样本中的位数. 试证明：

i) $E_\theta[|x^*_{[\frac{n}{2}]}|^{\frac{n-1}{2}}]<\infty$;

ii) $P(|x^*_{[\frac{n}{2}]}-\theta|>\varepsilon>0)$以 n 的指数速度趋于零;

iii) θ 的 MLE 具有 Bahadur 渐近有效性.

11. 在题 5 中，试证明 $\frac{1}{\theta}(\theta\neq 0)$、$\frac{\sigma}{\theta}$ 的 MLE 是 BAN 估计，但不具有 Cramér 渐近有效性. 如何改进它，使之既是 BAN 估计，又具有 Cramér 渐近有效性.

12. 在题 6 中，试证明 $\ln\lambda$ 的 MLE 不存在. 并构造 $\ln\lambda$ 的估计，使之既是 BAN 估计，又具有 Cramér 渐近有效性.

13* 若总体服从多参数指数族分布

$$\frac{dP_\theta}{d\mu}=f(x,\ \theta)=\beta(\theta)e^{\theta' x}.$$

其自然参数域 Ω 是开集. 试证明 θ 的 MLE 具有 Bahadur 渐近有效性.

参 考 文 献

[1] P. W. Zehna: Invariance of maximum likelihood estimation Ann. Math. Statist. 37: 755(1966).

[2] Barndorff-Nielsen, O: Information and Exponential Families In Statistical Theory (1977).

[3] R. H. Berk: Consistency and asymptotic normality of MLE'S for exponential models, Ann. Math. Statist. 43: 193～204(1972).

[4] S. J. Haberman: Maximum likelihood estimates in exponential response models, Ann. Statist. 5: 815～841 (1977).

[5] 成平: i) «指数族分布参数的 MLE 存在性与唯一性»，科学通报，数学、物理学、化学专辑，123～126 (1980).

ii) 《极大似然估计的 Bahadur 渐近有效性》, 数学学报, 23: 883～900(1980).
iii) 《论极大似然估计的 Cramér 渐近有效性》, 科学通报, 25: 1057～1060 (1980).

[6] A. Wald: Note on the cosistency of the maximum likelihood estimates, Ann. Math. Statist. 20: 595～601 (1949).

[7] L. Lecam: On the asymptotic thery of estimation and testing hypotheses, Proc. Third Berkeley Symp.Math. Statist.Prob. I 129～156 (1956).

[8] P. J. Huber: The behavior of maximum likelihood estimates under non-standard conditions, Proc. Fifth Berkeley symp. Math. Statist. Prob. I, 221～233 (1967).

[9] R. R. Bahadur: On the asymptotic efficiency of tests and estimators, Sankhya 22: 229～251 (1960).
Example of inconsistency of maximum likelihood estimates, Sankhya 20: 207～210 (1958).

[10] M. D. Perlman: On the strong consistency of approximate maximum likelihood estimators, Proc. Sixth Berkeley Symp. Math. Statist. prob. I, 63～282 (1972).

[11] J. Kiefer and J. Wolfowitz: Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters, Ann. Math. Statist. 27: 887～906 (1956).

[12] R. A. Fisher: Theory of statistical estimation, Proc. Camb. Phil. Soc. 20: 700～715 (1925).

[13] H. Cramér: Mathematical Method In Statistics, Princeton(1948).

[14] H. E. Danies: The asymptotic efficiency of a maximum likelihood estimator, Proc. Fourth Berkeley Symp. Math. Statist. Prob. I: 151～164(1961).

[15] J. Wolfowitz: Asymptotic efficiency of maximum likelihood estimators, Theory of Prob. and Appl. 10: 247～260 (1965).

[16] L. Weiss, J. Wolfowitz: i) Generalized maximum likelihood estimators, Theory of Prob. and Its Appl. 11: 58～81 (1966).
ii) Maximum probability estimators, Ann. Inst. Stat. Math. 19: 193～206 (1967).
iii) Maximum Probability Estimators And Relate Topic, Lecture Notes in Mathematics, No. 424 Springer-Verlag, Berlin-Heidelberg New York (1974).

[17] C. R. Rao: i) Asymptotic efficiency and limiting information, Proc. Fourth Berkeley Symp. Math. Statist. Prob. I: 531～546(1961).
ii) Efficiency estimates and optimum inference procedures in large samples, J. Roy. Stat. Soc. 2: 46～72 (1962).
iii) Criteria of estimation in large samples, Sankhya series A 25: 189～206

(1963).

iv) Linear Statistical Inference And Its Application, John Wiely New York (1965).

[18] Kai Takenchi and Masafumi Akahira: Lecture Notes in Mathematics Proc. Third Japan-USSR Symp. on Prob. Theory (1976).

[19] S. S. Peng: Rate of convergence of estimators based on sample mean, Ann. Statist. 6: 1048～1056 (1978).

[20] 卢昆亮:《截断参数估计的收敛速度》,科学通报,第1期, 11～13 (1982).

[21] C. P. Sidney and C. J. Stone: Fisher information and the Pitman estimator of a location parameter, Ann. Statist. 2: 225～247(1974).

[22] V. V. Petrov: Sums of Independent Random Variables (1975).

[23] W. Commins: Jr. Asymptotic variance as an approximation to expected loss for maximum likelihood estimates, Tech. Rep. No 26 Contrad N 60nr 23140(NR 342～022) Dept. of Statist. Stanford University (1959).

[24] D. Basu: An inconsistency of the method of maximum likelihood Ann. Math. Statist. 27: 336～351(1956).

[25] G. Kuldorf: Estimation From Grouped Data And Partically Samples, John Wiley New York (1959).

[26] 陈家鼎: i)《关于定时截止寿命试验的最大似然估计》,数学学报, 20: 145～148 (1977).

ii)《关于截尾样本情形下最大似然估计》,应用数学学报, 3: 306～321(1980).

第七章 同变估计

§1 同变性概论

我们通过下面几个例子，逐步引进同变的概念，并加以说明.

例1.1 比较两个不同的处理，它们的结果分别服从正态分布 $N(\mu_1, \sigma^2)$ 及 $N(\mu_2, \sigma^2)$. 每个处理分别得到 n 个结果: $x=(x_1, \cdots, x_n)^\tau$, $y=(y_1, \cdots, y_n)^\tau$. 要问两种处理是否有显著的差异. 具体说: 给定 $\Delta>0$, 令

$$\Theta_1=\{(\mu_1, \mu_2, \sigma^2): |\mu_1-\mu_2|\leqslant\Delta\};$$

$$\Theta_2=\Theta-\Theta_1=\{(\mu_1, \mu_2, \sigma^2): |\mu_1-\mu_2|>\Delta\}.$$

$\theta=(\mu_1, \mu_2, \sigma^2)$, 以 d_i 表示判断 $\theta\in\Theta_i (i=1, 2)$. 其损失函数当 $\theta\in\Theta_i$ 时，为 $L(\theta, d_i)=0$; 当 $\theta\overline{\in}\Theta_i$ 时，为 $L(\theta, d_i)=L>0$. 由于损失函数及 (x, y) 的分布都是 μ_1, μ_2 的对称函数，自然会要求判决函数 $d(x, y)$ 也是 x, y 的对称函数，即 $d(x, y)=d(y, x)$.

例1.2 研究两个随机变量 X、Y 的相关系数 ρ. 众所周知，$aX+b$, $aY+b$ 之间的相关系数和 X、Y 之间的相关系数相同. 如果已得 (X, Y) 的一组简单样本 (x, y), $x=(x_1, \cdots, x_n)^\tau$, $y=(y_1, \cdots, y_n)^\tau$. 要用样本估计相关系数时，自然想到在样本空间线性变换下 $(x, y)\to(ax+b\mathbb{1}, ay+b\mathbb{1})$, 此处 $\mathbb{1}=(1, \cdots, 1)^\tau$, ρ 的估计量应保持不变.

例1.3 废品率 p 的估计. 随机变量 X 仅取值 0、1, 0 值表示合格品; 1 表示废品. 现在随机独立地检查了 n 个产品，结果为 $x=(x_1, \cdots, x_n)^\tau$. 若废品率 p 的估计为 $d(x)$, 估计的损失函数为 $(p-d)^2$. 注意到可用变换 $x_i\to1-x_i$, $p\to1-p$, $d\to1-d$, 把估计废品率问题换为估计合格品率. 这是相互等价的问题，损失

函数不变，自然希望估计量具有相应的性质，即 $d(x)=1-d(1-x)$.

例 1.4 设随机变量 X 服从分布 $F((x-\mu)/\sigma)$，此处 $-\infty<\mu<\infty$, $0<\sigma<\infty$. μ 称为位置参数，若 $\int x dF(x)=0$, μ 就表示均值；σ 称为刻度参数，是离散度的标志．现要估计 μ, 损失函数为 $(d-\mu)^2/\sigma^2$. 注意，若 x 取不同刻度(单位)，μ 和 σ 也应与 x 有相同的刻度(单位)．用数学语言说，就是当 $x\to ax$, $a>0$, 则 $\mu\to a\mu$, $\sigma\to a\sigma$. 如果 $d\to ad$, 则损失函数不变：$(d-\mu)^2/\sigma^2=(ad-a\mu)^2/a^2\sigma^2$. 我们自然希望 μ 的估计量具有这种同变性质，即 $d(x)=d(ax)/a$.

从上面四个例子，可以找到某些共性：

1．对样本空间作某种变换，所有类似的变换构成一个群，记为 $\mathscr{G}$. 在例 1 和例 3 中变换群 $\mathscr{G}$ 是有限群，包含两个元素 (g_0, g_1). 在例 1 中，$g_0(x, y)=(x, y)$, $g_1(x, y)=(y, x)$; 在例 3 中，$g_0x=x$, $g_1x=1-x$, $g_0=g_0^2=g_1^2$ 为单位元；在例 2 中，变换群是线性变换群(仿射群)；在例 4 中，是乘法群．$\mathscr{G}$ 的每个元素 g 作用在样本空间上，引起样本空间的自身变换，但样本分布类型并不起变化，只是参数起变化．由此产生参数空间变换群 $\overline{\mathscr{G}}$, 它与 $\mathscr{G}$ 同态.

2．如果行动空间也作相应的变换，这些变换也构成一个群 $\mathscr{G}^*$, 它与 $\mathscr{G}$ 同态(不必同构，例如在例 1 和例 2 中，$\mathscr{G}^*$ 仅有单位元)．此外，参数空间、行动空间分别在 $g\in\mathscr{G}$ 所对应的 $\bar{g}$ 和 g^* 作用下，损失函数的值仍不变.

3．注意到上面 1、2 所述的同变特性，自然期望在作判决函数时，估计量也具有同变性质，即 $d(gx)=g^*d(x)$, $\forall g\in\mathscr{G}$.

现在我们给出严格的数学描述：

定义 1.1 不变统计判决问题：设 $(\mathscr{X}, \mathscr{B}_{\mathscr{X}})$ 是随机变量 X 的样本空间，$\mathscr{P}=\{P_\theta, \theta\in\Theta\}$ 为样本空间上的概率分布族(规定当 $\theta_1\neq\theta_2$ 时，$P_{\theta_1}\neq P_{\theta_2}$).

1．令 $\mathscr{G}$ 为 $\mathscr{X}\to\mathscr{X}$ 一一对应的可测变换群，即对 $\forall g\in\mathscr{G}$, 当

$c\in\mathscr{B}_{\mathscr{X}}$，有 $gc\in\mathscr{B}_{\mathscr{X}}$（对于群 $\mathscr{G}$ 来说，这等价于 $g\mathscr{B}_{\mathscr{X}}=\mathscr{B}_{\mathscr{X}}$，$\forall g\in\mathscr{G}$）.

2．若 $X\sim P_\theta$，假定对 $\forall g\in\mathscr{G}$，存在 $\theta'\in\Theta$（θ' 与 g 有关），使 $gX\sim P_{\theta'}$．由此定义，$\bar{g}$ 为 $\theta'=\bar{g}\theta$，它是 $\Theta\to\Theta$ 的变换，由此导出 $\bar{g}=\{\bar{g},\ g\in\mathscr{G}\}$（可以证明 $\overline{\mathscr{G}}$ 构成与 $\mathscr{G}$ 同态的群）.

3．给定了行动空间 $(D,\ \mathscr{B}_D)$ 及损失函数 $L(\theta,\ d)\geqslant 0$，在固定 θ 时，$L(\theta,\ d)$ 关于 $\mathscr{B}_D$ 可测[若考虑参数空间有可测结构 $(\Theta,\ \mathscr{B}_\Theta)$，则要求 $L(\theta,\ d)$ 关于 $(\Theta\times D,\ \mathscr{B}_\Theta\times\mathscr{B}_D)$ 可测]．我们假定对任何 $g\in\mathscr{G}$，存在一个由 $D\to D$ 的变换 g^*，使得

i) $L(\bar{g}\theta,\ g^*d)=L(\theta,\ d)$ 对 $\forall\theta\in\Theta,\ d\in D$. (1.1)

ii) g^* 为 $D—D$ 的一一对应变换，且 $g^*\mathscr{B}_D=\mathscr{B}_D$.

iii) $\mathscr{G}^*=\{g^*,\ g\in\mathscr{G}\}$ 是一个与 $\mathscr{G}$ 同态的群.

在上述条件下，则称以 $(\mathscr{X},\ \mathscr{B}_{\mathscr{X}},\ \mathscr{P})$ 和 $L(\theta,\ d)$ 为标志的统计判决问题在群 $\mathscr{G}$ 之下保持不变，有时简称为**不变统计判决问题**.

本章讨论不变统计判决问题时，在涉及参数空间先验分布时，我们皆假设 $\bar{g}\mathscr{B}_\Theta=\mathscr{B}_\Theta$ 对 $\forall\bar{g}\in\overline{\mathscr{G}}$ 成立.

在假设检验时，D 与 Θ 差别很大，而对参数估计，常常 $(\Theta,\ \mathscr{B}_\Theta)$ 与 $(D,\ \mathscr{B}_D)$ 等价，$\overline{\mathscr{G}}$ 与 $\mathscr{G}^*$ 有时同构，故时常不加区别地用相同符号.

定义 1.2 同变估计．在不变统计判决问题中，θ 的一个非随机化估计 $\hat{\theta}(x)$（是由 $(\mathscr{X},\ \mathscr{B}_{\mathscr{X}})$ 到 $(D,\ \mathscr{B}_D)$ 的可测函数）称之为**同变非随机化估计**，如果对 $\forall g\in\mathscr{G}$，$\forall x\in\mathscr{X}$，有

$$\hat{\theta}(gx)=g^*\hat{\theta}(x). \tag{1.2}$$

若(1.2)式仅仅几乎处处($\mathscr{P}$)成立（对每个 g，(1.2)不成立的 x 零测集可与 g 有关．当 $\mathscr{P}\ll\mu$ 时，往往是关于 μ 几乎处处成立），则称为**几乎同变非随机化估计**．有时简称为**同变估计**或**几乎同变估计**.

定义在 $\mathscr{X}\times\mathscr{B}_D$ 上的函数 $\delta(A|x)$，$A\in\mathscr{B}_D$，$x\in\mathscr{X}$ 是 θ 的随机化估计，在固定 x 时它是 $\mathscr{B}_D$ 上的概率测度，在固定 A 时，它是

$\mathscr{B}_{\mathscr{X}}$可测函数. 如果它还满足对 $\forall g\in\mathscr{G}$, $\forall x\in\mathscr{X}$, 有

$$\delta(g^*A|gx)=\delta(A|x) \tag{1.3}$$

成立, 则称 δ 为随机化同变估计. 如果上式仅要求几乎处处 ($\mathscr{P}$) 成立, 即对每个 $g\in\mathscr{G}$, 存在 $\mathscr{P}$ 零测集 N_g, 仅当 $x\overline{\in}N_g$ 时(1.3)式成立, 则称 δ 为随机化的几乎同变估计.

显然, 随机化估计类包含了非随机化估计类.

定义 1.3 不变可测集、不变 σ–域、最大不变 σ–域、不变统计量、最大不变统计量. 对某个 $B\in\mathscr{B}_{\mathscr{X}}$, 若对 $\forall g\in\mathscr{G}$ 有 $gB=B$, 则称 B 为关于 $\mathscr{G}$ 的不变可测集; 易见, 所有不变可测集组成 $\mathscr{B}_{\mathscr{X}}$ 的一个子 σ–域, 称为最大不变 σ–域; 它的任意子 σ–域称为不变 σ–域. 如果由统计量 $T(x)$ 生成的 σ–域是不变 σ–域, 则称 $T(x)$ 为不变统计量; 如果由 $T(x)$ 生成的 σ–域是最大不变 σ–域, 则称 $T(x)$ 为最大不变统计量, 显然, 它会有多种不同的表现形式.

对任给的 $x_0\in\mathscr{X}$, 集合$\{x:gx=x_0,\ g\in\mathscr{G}\}$称为 $\mathscr{G}$ 变换下的一条样本轨道.

引理 1.1 若 $T(x)$ 为样本空间$(\mathscr{X},\ \mathscr{B}_{\mathscr{X}})$上的统计量, $T(x)$ 的值域$(\mathscr{T},\ \mathscr{B}_{\mathscr{T}})$为欧氏空间, $\mathscr{B}_{\mathscr{T}}$ 由 $\mathscr{T}$ 的 Borel 子集所组成. 则

i) $T(x)$是不变统计量$\Leftrightarrow T(gx)=T(x)$, $\forall x\in\mathscr{X}$, $\forall g\in\mathscr{G}$.

ii) $T(x)$是最大不变统计量$\Leftrightarrow T(x)$不变, 且对任意的不变统计量 s, 皆有 $s=\varphi(T)$, φ 为$(\mathscr{T},\ \mathscr{B}_{\mathscr{T}})\to(\mathscr{R}_1,\ \mathscr{B}_1)$的可测函数.

证明 i)的必要性: 因为 $T(x)$是不变统计量, 故对$\forall x_0\in\mathscr{X}$, $\{x:T(x)=T(x_0)\}$是不变可测集, 于是对于$\forall g\in\mathscr{G}$, $gx_0\in\{x:T(x)=T(x_0)\}$, 这就是说 $T(gx_0)=T(x_0)$.

i)的充分性: 只要证对于 $\forall C\in\mathscr{B}_{\mathscr{T}}$, $T^{-1}(C)$ 是不变集即可. 对于 $\forall g\in\mathscr{G}$, 有

$$\begin{aligned}gT^{-1}(C)&=g\{x:T(x)\in C\}=g\{x:T(gx)\in C\}\\&=\{gx:T(gx)\in C\}=\{x:T(x)\in C\}=T^{-1}(C).\end{aligned}$$

ii) 的必要性: 设 $T(x)$ 为最大不变统计量, 即 $\mathscr{B}(T)$ 为最大不变 σ–域. 则对任意不变的 $s(x)$, 应有 $\mathscr{B}(s)\subset\mathscr{B}(T)$, 由可测函数的复合函数定理(见第一章定理 4.1), 则存在 φ, 使得 $s=\varphi(T)$.

ii)的充分性：对任意不变可测集 A，I_A 是不变统计量，故 $I_A=\varphi(T)$，从而知 $A\in\mathscr{B}(T)$，引理证毕.

引理 1.2 如果对任意的 $\theta_1\neq\theta_2$，$P_{\theta_1}\neq P_{\theta_2}$，$\mathscr{G}$ 是 $(\mathscr{X},\mathscr{B}_{\mathscr{X}})$ 上的可测变换群，使分布族保持不变，则由它诱导出参数空间 $\overline{\mathscr{G}}$ 构成一个群与 g 同态. 而且对 $\forall\bar{g}\in\overline{\mathscr{G}}$，有 $\bar{g}\Theta=\Theta$.

证明 设 $X\sim P_\theta$，对 $\forall g\in\mathscr{G}$，由假设，$\mathscr{G}$ 使分布族保持不变，故存在 $\theta'\in\Theta$，使 $gX\sim P_{\theta'}$，且由 $\theta_1\neq\theta_2\Rightarrow P_{\theta_1}\neq P_{\theta_2}$ 知 θ' 是唯一的，故导出 $\bar{g}:\bar{g}\theta=\theta'$，且可知 $\bar{g}\Theta=\Theta$，对于 $\forall g_1、g_2\in\mathscr{G}$，有 $g=g_1g_2$，在 $\overline{\mathscr{G}}$ 中的相应元素为 $\bar{g}_1$，$\bar{g}_2$，$\bar{g}_3$，由于 $gX\sim P_{\theta'}$，故对于 $\forall A\in\mathscr{B}_{\mathscr{X}}$，有

$$
\begin{aligned}
P_{\theta'}(A)&=P_\theta(g^{-1}A)=P_\theta(g_2^{-1}g_1^{-1}A)\\
&=P_{\bar{g}_2\theta}(g_1^{-1}A)=P_{\bar{g}_1\cdot\bar{g}_2\theta}(A).
\end{aligned}
$$

由于 $\theta_1\neq\theta_2$ 时，$P_{\theta_1}\neq P_{\theta_2}$，所以 $\bar{g}_1\cdot\bar{g}_2\theta=\theta'=\bar{g}\theta$. 从而知 $\bar{g}=\overline{g_1g_2}=\bar{g}_1\cdot\bar{g}_2$. 特别当 $g_2=g_1^{-1}$ 时，$\overline{g_1g_2}$ 为 $\overline{\mathscr{G}}$ 中的单位元 $\bar{I}$($\bar{I}\theta=\theta$，对于 $\forall\theta\in\Theta$)，$\bar{g}_1\cdot\bar{g}_1^{-1}=\bar{I}$，即 $\bar{g}_1^{-1}=(\bar{g}_1)^{-1}$. 这就说明 $\overline{\mathscr{G}}$ 构成一个群，而且与 $\mathscr{G}$ 同态. 引理证毕.

对于变换群 $\mathscr{G}$，除在样本空间给出轨道外，还在参数空间给出轨道：对于任意的 $\theta_0\in\Theta$，$\{\theta:\bar{g}\theta=\theta_0,\ \bar{g}\in\overline{\mathscr{G}}\}$ 称为 Θ 的一条轨道. 显然，Θ 的所有点可按是否属于同一轨道而划分为等价类. 令 $\{\Theta_s,\ s\in S\}$ 表示 Θ 在 $\overline{\mathscr{G}}$ 变换下的所有轨道，S 为等价类集. 下面我们给出同变估计的一个特征：同变估计的风险函数在 Θ 的同一轨道 Θ_s 上取相同的值.

定理 1.1 以 $(\mathscr{X},\mathscr{B}_{\mathscr{X}},\mathscr{P})$，$\mathscr{P}=\{P_\theta,\ \theta\in\Theta\}$ 和 $L(\theta,d)$ 为标志的统计判决问题在群 $\mathscr{G}$ 下保持不变，则对任何 θ 的同变估计 δ(随机化或非随机化)，它的风险函数 $R(\theta,\delta)$ 在 Θ 的同一轨道上取相同的值.

证明 只须考虑随机化的估计(易知非随机化的估计是它的特殊情形).

设 $\delta(a|x)$ 是 θ 的随机化同变估计，所以对于 $\forall g\in\mathscr{G}$，$\delta(g^*a|gx)=\delta(a|x)$，即 $\delta(dg^*a|gx)=\delta(da|x)$. 这里 d 是微分符号，所以

$$R(\theta, \delta) = E_\theta\left[\int_D L(\theta, a)\delta(da|x)\right]$$

$$= E_\theta\left[\int_D L(\bar{g}\theta, g^*a)\delta(dg^*a|gx)\right]$$

$$= E_\theta\left[\int_D L(\bar{g}\theta, a')\delta(da'|gx)\right]$$

$$= E_{\bar{g}\theta}\left[\int_D L(\bar{g}\theta, a')\delta(da'|x)\right] = R(\bar{g}\theta, \delta)$$

对 $\forall \bar{g} \in \bar{\mathscr{G}}$ 成立，从而定理得证.

§2 平方损失下某些最优同变估计

研究同变估计理论时，我们自然希望找到最好的同变估计，这就要规定最优的法则.

在不变统计判决问题中，同变估计 δ_0 称为**一致最优同变估计**，如果对于任何同变估计 δ，皆有下式成立：

$$R(\theta, \delta_0) \leqslant R(\theta, \delta), \quad 对于\ \forall \theta \in \Theta. \tag{2.1}$$

此处 $R(\theta, \delta)$ 为采用估计 δ 的风险函数.

在本节，我们要在平方损失下，探讨最优同变估计的求法和具体表达式，由于这是凸损失，我们在非随机化同变估计类中寻优即可.

一、转移分布族的位置参数的最优同变估计

设 k 维随机向量 $X \sim P(x - \theta\mathbb{1})$，这里 $\mathbb{1} = (1, \cdots, 1)'$，$-\infty < \theta < +\infty$. 损失函数 $L(\theta, d) = (\theta - d)^2$. 变换群 $\mathscr{G} = \{g: g \in R'\}$ 为加法实数群(平移群)：

$$gx = x + g\mathbb{1},\ \forall x \in R^k,\ \forall g \in \mathscr{G}.$$

由此群产生作用于参数空间的变换群 $\bar{\mathscr{G}} = \mathscr{G}$：

$$\bar{g}\theta = \theta + g,\ \forall g \in \mathscr{G},\ \forall \theta \in \Theta = \{\theta: -\infty < \theta < +\infty\}.$$

不难验证在参数空间只有唯一的轨道 Θ. 再若采用行动空间 D

的变换群 $\mathscr{G}^*=\mathscr{G}:g^*d=d+g$. 则这个统计判决问题不变.

设 $h(x)$ 是任意不变统计量, 根据定义及引理1.1, 当取 $g=-x_1$ 时,则有

$$h(x)=h(gx)=h(x-x_1\mathbb{1})=\psi(T(x)).$$

这里 $T(x)=(x_2-x_1, \cdots, x_k-x_1)$, 可见 $T(x)$是最大不变统计量. 当然, 表示法不必唯一. 对 θ 的任意同变估计 $d(x)$, 即 $\forall g\in\mathscr{G}$, $d(gx)=g^*d(x)=d(x)+g$, 取 $g=-x_1$, 有

$$d(x)=x_1+d(x-x_1\mathbb{1})=x_1+\psi(T(x)).$$

这就是平移群下同变估计的一般表现形式.

现在求最优同变估计, 当然得假设存在一个同变估计 $x_1+\varphi_0(T)$, 使 $E_0\|x_1+\varphi_0(T)\|^2<\infty$, 我们先设 $E_0\|X_1\|^2<\infty$, 故 $d(x)$ 的风险函数是

$$\begin{aligned}R(\theta, d(x))&=E_\theta(x_1+\psi(T(x))-\theta)^2\\&=E_0(x_1+\psi(T))^2=E_0\{E_0[(x_1+\psi(T))^2|T]\}\\&=E_0\{E_0[(X_1-E_0(X_1|T))^2|T]\}\\&\quad+E_0\{[E_0(X_1|T)+\psi(T)]^2\}\\&\quad+2E_0\{E_0[(X_1-E_0(X_1|T))(E_0(X_1|T)\\&\quad+\psi(T))|T]\}.\end{aligned}$$

注意,上式第三项为0, 故当 $\psi(T)=-E_0(X_1|T)$ 时, 使 $R(\theta, \delta)$ 达到最小值,也即最优同变估计形为

$$\delta_0(x)=x_1-E_0(X_1|T). \tag{2.2}$$

由损失函数的严凸性即可得最优同变估计的唯一性.

如果分布函数 $P(x-\theta\mathbb{1})$ 有密度函数 $p(x-\theta\mathbb{1})$, 作变换 $Y_1=X_1$, $Y_i=X_i-X_1(2\leqslant i\leqslant k)$, 则 $Y=(Y_1, \cdots, Y_k)'$ 在 $\theta=0$ 时的联合密度为 $p(y_1, y_2+y_1, \cdots, y_k+y_1)$,它在给定 $T=(y_2, \cdots, y_k)$ 的条件概率为

$$p(y_1, y_2+y_1, \cdots, y_k+y_1)\Big/\int_{-\infty}^{\infty}p(t, y_2+t, \cdots, y_k+t)dt,$$

故

$$E_0(X_1|T)=\frac{\int_{-\infty}^{\infty} y_1 p(y_1,\ y_2+y_1,\ \cdots,\ y_k+y_1)dy_1}{\int_{-\infty}^{\infty} p(y_1,\ y_2+y_1,\ \cdots,\ y_k+y_1)dy_1}$$

$$=\frac{\int_{-\infty}^{\infty}(x_1-\theta)p(x_1-\theta,\ y_2+x_1-\theta,\ \cdots,\ y_k+x_1-\theta)d\theta}{\int_{-\infty}^{\infty}p(x_1-\theta,\ y_2+x_1-\theta,\ \cdots,\ y_k+x_1-\theta)d\theta}$$

$$=x_1-\int_{-\infty}^{\infty}\theta p(x-\theta\mathbb{1})d\theta\Big/\int_{-\infty}^{\infty}p(x-\theta\mathbb{1})d\theta. \qquad (2.3)$$

其中第二步是令 $y_1=x_1-\theta_0$, 由 (2.3) 和 (2.2) 即得 θ 的最优同变估计的明显表达式为

$$\delta_0(x)=\int_{-\infty}^{\infty}\theta p(x-\theta\mathbb{1})d\theta\Big/\int_{-\infty}^{\infty}p(x-\theta\mathbb{1})d\theta. \qquad (2.4)$$

此形的估计一般称为 Pitman 估计.

如果 $E_0\|x_1\|^2=\infty$, 则把 X_1 换成 $X_1+\varphi_0(T)$, 用同样方法可以得到

$$\delta_0(x)=x_1+\varphi_0(T)-E_0(X_1+\varphi_0(T)\,|T).$$

把(2.3)的 $E_0(X_1|T)$ 换为 $E_0(X_1+\varphi_0(T)\,|T)$, 可算得它等于

$$x_1+\varphi_0(T)-\int_{-\infty}^{\infty}\theta p(x-\theta\mathbb{1})d\theta\Big/\int_{-\infty}^{\infty}p(x-\theta\mathbb{1})d\theta.$$

从而仍可得(2.4)式的结论.

很易证明 Pitman 估计是无偏估计. 若随机变量 $X\sim F(x-\theta)$ 有密度 $f(x-\theta)$, 随机独立抽样结果为 $x=(x_1,\ \cdots,\ x_n)^\tau$ 有联合密度为 $p(x-\theta\mathbb{1})=\prod_{i=1}^{n}f(x_i-\theta)$, 则 θ 的最优同变估计可表为

$$\delta_0(x)=\int_{-\infty}^{\infty}\theta\prod_{i=1}^{n}f(x_i-\theta)d\theta\Big/\int_{-\infty}^{\infty}\prod_{i=1}^{n}f(x_i-\theta)d\theta. \qquad (2.5)$$

例 2.1 令 X 表示寿命服从二参数指数分布, 其密度为

$$f(x-\alpha)=\beta^{-1}\exp[-(x-\alpha)/\beta]I_{(x\geq\alpha)}\quad(\alpha\in R_1,\ \beta>0).$$

今进行了 n 个独立试验, 获得了已结束寿命的 r 个数据 $x_{(1)}\leqslant\cdots\leqslant x_{(r)}$, 可知其联合密度为

$$\frac{n!}{(n-r)!}\beta^{-r}\exp\left[-\left(\sum_{i=1}^{r}x_{(i)}+(n-r)x_{(r)}-n\alpha\right)/\beta\right]I_{(x_{(1)}\geq\alpha)}.$$

记 $U=\sum_{i=1}^{r}x_{(i)}+(n-r)x_{(r)}$.

对已知的 β, 参数空间为 $\alpha\in R_1$, 在平方损失下, α 的最优同变估计 $\hat{\theta}(x_{(1)},\cdots,x_{(r)})$ 为

$$\frac{\int_{-\infty}^{x_{(1)}}\alpha\exp[-(U-n\alpha)/\beta]d\alpha}{\int_{-\infty}^{x_{(1)}}\exp[-(U-n\alpha)/\beta]d\alpha}$$

$$=\int_{-\infty}^{x_{(1)}}\alpha\exp[n\alpha/\beta]d\alpha\Big/\int_{-\infty}^{x_{(1)}}\exp[n\alpha/\beta]d\alpha$$

$$=x_{(1)}-\beta/n.$$

不难验算其方差为 β^2/n^2,与 r 无关.

例 2.2 令 $X\sim N(\mu,1)$ $(-\infty<\mu<\infty)$, $x_1,\cdots,x_n$ 为来自 X 的简单抽样结果, μ 的同变估计为

$$\frac{\int_{-\infty}^{\infty}\mu\exp\left[-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right]d\mu}{\int_{-\infty}^{\infty}\exp\left[-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right]d\mu}=\sum_{i=1}^{n}x_i/n=\bar{x}.$$

上述结果很易推广到多参数情形. 若 $k\times n$ 随机矩阵 $X\sim P(x-\theta 1^{\tau})$, 此处 $\theta\in R^k$, 则在损失函数 $L(\theta,d)=\|\theta-d\|^2$ 下, θ 的最优同变估计为

$$\hat{\theta}(x)=x_1-E_0(x_1|T). \tag{2.6}$$

此处 $T=(x_2-x_1,\cdots,x_n-x_1)$ 是随机矩阵.

当 $P(x-\theta 1^{\tau})$有密度函数 $p(x-\theta 1^{\tau})$时, 则 θ 的最优同变估计可表为

$$\hat{\theta}(x)=\int_{R^k}\theta p(x-\theta 1^{\tau})d\theta\Big/\int_{R^k}p(x-\theta 1^{\tau})d\theta. \tag{2.7}$$

例 2.3 设 $X\sim N(\mu,A_k)$, $A_k>0$ 已知, $\mu\in R^k$ 未知; 估计为 μ, 损失函数为 $\|\mu-d\|^2$. 对于来自 $N(\mu,A_k)$ 的 n 个独立观测 $x_1,\cdots,x_n$, μ 的最优同变估计为

$$\hat{\mu}=\frac{\int_{R^k}\mu\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^{\tau}A_k^{-1}(x_i-\mu)\right\}d\mu}{\int_{R^k}\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^{\tau}A_k^{-1}(x_i-\mu)\right\}d\mu}$$

$$=\frac{\int_{R^k}\mu\exp\left\{-\frac{n}{2}(\bar{x}-\mu)^\tau A_k^{-1}(\bar{x}-\mu)\right\}d\mu}{\int_{R^k}\exp\left\{-\frac{n}{2}(\bar{x}-\mu)^\tau A_k^{-1}(\bar{x}-\mu)\right\}d\mu}=\bar{x}.$$

它与 A_k 取何值无关，推导中利用了等式

$$\sum_{i=1}^{n}(x_i-\mu)^\tau A_k^{-1}(x_i-\mu)=\sum_{i=1}^{n}(x_i-\bar{x})A_k^{-1}(x_i-\bar{x})$$
$$+n(\bar{x}-\mu)^\tau A_k^{-1}(\bar{x}-\mu).$$

二、刻度参数分布族中刻度参数的最优同变估计

设 k 维随机向量 $X\sim P(x/\sigma)$，并假设 $P_0(\|X\|=0)=0$，变换群 $\mathscr{G}=\{g:g>0\}$ 取为一般乘法下的乘法群. 对于 $\forall g\in\mathscr{G}$, $gx=g\cdot x$, 由 $\mathscr{G}$ 引起参数空间 $\Theta=\{\sigma:\sigma>0\}$ 的变换群 $\bar{\mathscr{G}}=\mathscr{G}$, $\forall\bar{g}\in\bar{\mathscr{G}}$, 有 $\bar{g}\sigma=\bar{g}\cdot\sigma=g\sigma$. 估计 σ 的损失函数取为 $L(\sigma, d)=(\sigma-d)^2/\sigma^2$. 那末行动空间的变换群采用 $\mathscr{G}^*=\mathscr{G}$: $\forall g^*\in\mathscr{G}^*$, $g^*d=gd$, 就使判决问题不变.

对任一同变估计 $d(x)$, 若取 $g^*=\|x\|^{-1}$, 可得

$$d(x)=g^{*-1}d(gx)=\|x\|d(\|x\|^{-1}x)=\|x\|d(T).$$

此处 $T=\|x\|^{-1}x$, 事实上，它是最大不变统计量.

同前一问题一样，要假设有同变估计在 $\sigma=1$ 时均方有限，不失一般性，可先设 $E_1\|x\|^2<\infty$. 由于

$$\begin{aligned}E_\sigma(d(x)-\sigma)^2/\sigma^2&=E_\sigma(\|x\|d(T)-\sigma)^2/\sigma^2\\&=E_1(\|x\|d(T)-1)^2=E_1\{E[(\|x\|d(T)-1)^2|T]\}\\&\geqslant E_1\{\min_d[d^2(T)E_1(\|x\|^2|T)-2d(T)E_1(\|x\||T)+1]\}\\&=E_1(\|x\|E_1(\|x\||T)/E_1(\|x\|^2|T)-1)^2.\end{aligned}$$

由此得到 σ 的最优同变估计为

$$\hat{\sigma}=\|x\|E_1(\|x\||T)/E_1(\|x\|^2|T). \tag{2.8}$$

令 $Y_1=\|x\|$, $Y_i=x_i/\|x\|(2\leqslant i\leqslant k)$. 注意到它的 Jacobian 行列式值为 $y_1^{k-1}\Big/\left(1-\sum_{i=2}^{k}y_i^2\right)^{1/2}$. 于是当 $P(x/\sigma)$ 有密度函数

$\sigma^{-k}p(x/\sigma)$时，$\sigma=1$时$(y_1\cdots y_k)$的联合密度为

$$y_1^{k-1}\left(1-\sum_{i=2}^k y_i^2\right)^{-1/2}\left[p\left(y_1\left(1-\sum_{i=2}^k y_i^2\right)^{1/2},\ y_1y_2,\ \cdots,\ y_1y_k\right)\right.$$

$$\left.+p\left(-y_1\left(1-\sum_{i=2}^k y_i^2\right)^{1/2},\ y_1y_2,\ \cdots,\ y_1y_k\right)\right].$$

注意，T 正好与$(\text{Sgn}\,x_1,\ y_2,\ \cdots,\ y_k)$等价，故

$$\frac{E_1(\|x\|\,|T)}{E_1(\|x\|^2|T)}=\left\{\int_0^\infty y_1^k\left(1-\sum_{i=2}^k y_i^2\right)^{-1/2}p\left(y_1\left(1-\sum_{i=2}^k y_i^2\right)^{1/2}\right.\right.$$

$$\left.\left.\times\text{Sgn}\,x_1,\ y_1y_2,\ \cdots,\ y_1y_k\right)dy_1\right\}\Big/$$

$$\left\{\int_0^\infty y_1^{k+1}\left(1-\sum_{i=2}^k y_i^2\right)^{-1/2}p\left(y_1\left(1-\sum_{i=2}^k y_i^2\right)^{1/2}\right.\right.$$

$$\left.\left.\times\text{Sgn}\,x_1,\ y_1y_2,\ \cdots,\ y_1y_k\right)dy_1\right\}$$

对积分变量 y_1 作变量 $y_1\left(1-\sum_{i=2}^k y_i^2\right)^{1/2}\text{Sgn}\,x_1=x_1/\sigma$，再将 $y_i=x_i/\|x\|\,(2\leqslant i\leqslant k)$ 代入表达式，从而推得 $y_1=\|x\|/\sigma$，有

$$\frac{E_1(\|x\|\,|T)}{E_1(\|x\|^2|T)}=\frac{\int_0^\infty\sigma^{-k-2}p(x/\sigma)d\sigma}{\|x\|\int_0^\infty\sigma^{-k-3}p(x/\sigma)d\sigma}.$$

由此得到σ的最优同变估计为

$$\hat{\sigma}=\int_0^\infty\sigma^{-k-2}p(x/\sigma)d\sigma\Big/\int_0^\infty\sigma^{-k-3}p(x/\sigma)d\sigma.\qquad(2.9)$$

例 2.4 问题如例 2.1，参数 α 已知，而要估计刻度参数 β、$\Theta=\{\beta:\beta>0\}$、损失函数为$(\beta-d)^2/\beta^2$. 显然，此判决问题是在刻度变换群下的不变判决问题. β 的最优同变估计 $\hat{\beta}(x_{(1)},\ \cdots,\ x_{(r)})$ 应为

$$\hat{\beta}=\frac{\int_0^\infty\beta^{-r-2}\exp(-(U-n\alpha)/\beta)d\beta}{\int_0^\infty\beta^{-r-3}\exp(-(U-n\alpha)/\beta)d\beta}$$

$$=\int_0^\infty z^re^{-z(U-n\alpha)}dz\Big/\int_0^\infty z^{r+1}e^{-z(U-n\alpha)}dz$$

$$=(U-n\alpha)\Gamma(r+1)/\Gamma(r+2)$$

$$=\left(\sum_{i=1}^{r}x_{(i)}+(n-r)x_{(r)}-n\alpha\right)\Big/(r+1).$$

在所设分布之下，可知 $\alpha(U-n\alpha)/\beta\sim\chi^2_{2r}$，因此估计的风险函数 $E(\hat{\beta}-\beta)^2/\beta^2=1/(r+1)$.

三、刻度、位置参数皆未知时的最优同变估计

设 k 维随机向量 $X\sim P((x-\theta\mathbb{1})/\sigma)$，$P$ 具有密度函数 $\sigma^{-k}p((x-\theta\mathbb{1})/\sigma)\ (-\infty<\theta<\infty,\sigma>0)$. 在平方损失 $a(d_1-\theta)^2/\sigma^2+b(d_2-\sigma)^2/\sigma^2$ 下($a\geqslant0$, $b\geqslant0$ 已知,不同时为 0).

变换群为仿射群

$$\mathscr{G}=\left(\begin{pmatrix}1, & 0\\ e, & c\end{pmatrix}, -\infty<e<+\infty, c>0\right), \forall g\in\mathscr{G},$$

$$gx=cx+e\mathbb{1}.$$

$\mathscr{G}$是由平移群与乘法群结合而成的，在矩阵运算下，它是一个群. 也就是说,对于

$$g_1=\begin{pmatrix}1 & 0\\ e_1 & c_1\end{pmatrix},\ g_2=\begin{pmatrix}1 & 0\\ e_2 & c_2\end{pmatrix},\ g=g_1g_2\triangleq\begin{pmatrix}1 & 0\\ e_1 & c_1\end{pmatrix}\begin{pmatrix}1 & 0\\ e_2 & c_2\end{pmatrix}$$
$$=\begin{pmatrix}1 & 0\\ e_1+c_1e_2 & c_1c_2\end{pmatrix}.$$

而这可以对应着

$$g_1g_2x=g_1(c_2x+e_2\mathbb{1})=c_1(c_2x+e_2\mathbb{1})+e_1\mathbb{1}=c_1c_2x+(c_1e_2+e_1)\mathbb{1}.$$

由 $\mathscr{G}$ 诱导的参数空间的变换群 $\overline{\mathscr{G}}=\mathscr{G}:\forall\bar{g}\in\overline{\mathscr{G}}$:

$$\bar{g}(\theta, \sigma)^\tau=(c\theta+e, c\sigma)^\tau.$$

若行动空间变换群 $\mathscr{G}^*$ 也采用 $\mathscr{G}:\forall g^*\in\mathscr{G}^*$,

$$g^*d=(cd_1+e, cd_2)^\tau.$$

显然, $\mathscr{G}$, $\overline{\mathscr{G}}$, $\mathscr{G}^*$ 同构,并使判决问题不变.

对于$(\theta, \sigma)^\tau$ 的任一同变估计 $d(x)=(d_1(x), d_2(x))^\tau$,由定义知，$d(x)=g^{*-1}d(gx)$，对 $\forall g\in\mathscr{G}$ 成立. 现取 $g=\begin{pmatrix}1 & 0\\ x_1 & \tilde{s}\end{pmatrix}^{-1}$，这里 $\tilde{s}=\left(\sum_{i=2}^{k}(x_i-x_1)^2\right)^{1/2}$,则

$$d(x)=g^{*-1}d\left(\frac{x-x_1\mathbb{1}}{\tilde{s}}\right)=\left(x_1+d_1\left(\frac{x-x_1\mathbb{1}}{\tilde{s}}\right)\tilde{s},\ \tilde{s}d_2\left(\frac{x-x_1\mathbb{1}}{\tilde{s}}\right)\right)^\tau.$$

令 $T=(x-x_1\mathbb{1})/\tilde{s}$，它是最大不变统计量，一般最优同变估计就可表为

$$d(x)=(x_1+d_1(T)\tilde{s},\ \tilde{s}d_2(T))^\tau.$$

如同在本节第一小节中那样，不失一般性，我们假设 $E\|x\|^2<\infty$，则对任一同变估计 d，有

$$\begin{aligned}R(\theta,\sigma;d)&=aE_{\sigma,\theta}(d_1(x)-\theta)^2/\sigma^2+bE_{\sigma,\theta}(d_2(x)-\sigma)^2/\sigma^2\\&=aE_{1,0}(d_1(x))^2+bE_{1,0}(d_2(x)-1)^2\\&=aE_{1,0}(x_1+d_1(T)\tilde{s})^2+bE_{1,0}(\tilde{s}d_2(T)-1)^2\\&=aE_{1,0}\{E_{1,0}[(x_1+d_1(T)\tilde{s})^2|T]\}\\&\quad+bE_{1,0}\{E_{1,0}[(\tilde{s}d_2(T)-1)^2|T]\}\\&\geqslant aE_{1,0}\{\min_{d_1(T)}E_{1,0}[(x_1+d_1(T)\tilde{s})^2|T]\}\\&\quad+bE_{1,0}\{\min_{d_2(T)}E_{1,0}[(\tilde{s}d_2(T)-1)^2|T]\}\\&=E_{1,0}\left\{aE_{1,0}\left[\left(x_1-\frac{\tilde{s}E_{1,0}(x_1\tilde{s}|T)}{E_{1,0}(\tilde{s}^2|T)}\right)^2\Big|T\right]\right.\\&\quad\left.+bE_{1,0}\left[\left(\tilde{s}\frac{E_{1,0}(\tilde{s}|T)}{E_{1,0}(\tilde{s}^2|T)}-1\right)^2\Big|T\right]\right\}\\&=aE_{1,0}\left\{x_1-\frac{E_{1,0}(x_1\tilde{s}|T)\tilde{s}}{E_{1,0}(\tilde{s}^2|T)}\right\}^2\\&\quad+bE_{1,0}\left\{\tilde{s}\frac{E_{1,0}(\tilde{s}|T)}{E_{1,0}(\tilde{s}^2|T)}-1\right\}^2\end{aligned}$$

由此就得到 θ、σ 的最优同变估计为：

$$\hat{\theta}=x_1-\tilde{s}E_{1,0}(x_1\tilde{s}|T)/E_{1,0}(\tilde{s}^2|T);\tag{2.10}$$

$$\hat{\sigma}=\tilde{s}E_{1,0}(\tilde{s}|T)/E_{1,0}(\tilde{s}^2|T).\tag{2.11}$$

问题转为求条件数学期望，以得到 $\hat{\theta}$ 及 $\hat{\sigma}$ 的明显表达式．这求法与刻度参数已知时类似．

作变换 $y_1=x_1$，$y_2=\tilde{s}$，$y_i=(x_i-x_1)/\tilde{s}\ (i=3,\cdots,k)$，则 $(y_1,y_2,\cdots,y_k,\mathrm{Sgn}\,(x_2-x_1))$ 的联合密度为

$$\frac{1}{\sigma^k} p\left(\frac{y_1-\theta}{\sigma}, \frac{\operatorname{Sgn}(x_2-x_1)\cdot y_2\left(1-\sum_{i=3}^k y_i^2\right)^{1/2}+y_1-\theta}{\sigma},\right.$$

$$\left.\frac{y_2y_3+y_1-\theta}{\sigma}, \cdots, \frac{y_ky_2+y_1-\theta}{\sigma}\right) J(y).$$

这里 J 为变换的 Jacobian 行列式，因为此变换等于先作变换 $y_1=x_1$, $z_i=x_i-x_1$ $(i=2, \cdots, k)$，然后再作变换 $y_1=y_1$, $y_2=\tilde{s}$, $y_i=z_i/\tilde{s}$ $(i=3, \cdots, k)$. 故 J 为

$$\left|\det\begin{pmatrix} 1 & 0 & \cdots & 0 \\ -1 & 1 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ -1 & 0 & \cdots & 1 \end{pmatrix}\right.$$

$$\left.\times\det\begin{pmatrix} \left(1-\sum_{i=3}^k y_i^2\right)^{1/2} & -y_3\left(1-\sum_{i=3}^k y_i^2\right)^{-1/2} & \cdots & -y_k\left(1-\sum_{i=3}^k y_i^2\right)^{-1/2} \\ y_3 & y_2 & & 0 \\ \vdots & \vdots & \ddots & \vdots \\ y_k & 0 & \cdots & y_2 \end{pmatrix}\right|$$

$$=y_2^{k-2}\left(1-\sum_{i=3}^k y_i^2\right)^{-1/2}.$$

因此

$$\begin{aligned} &E_{1,0}(x_1\tilde{s}\,|\,T)/E_{1,0}(\tilde{s}^2\,|\,T) \\ &=\left\{\int_0^\infty\int_{-\infty}^\infty y_1y_2 p\left(y_1, y_2\left(1-\sum_{i=3}^k y_i^2\right)^{1/2}\cdot\operatorname{Sgn}(x_2-x_1)\right.\right. \\ &\qquad\left.\left.+y_1, y_2y_3+y_1, \cdots, y_2y_k+y_1\right)J\,dy_1\,dy_2\right\}\Big/ \\ &\qquad\left\{\int_0^\infty\int_{-\infty}^\infty y_2^2 p\left(y_1, y_2\left(1-\sum_{i=3}^k y_i^2\right)^{1/2}\cdot\operatorname{Sgn}(x_2-x_1)\right.\right. \\ &\qquad\left.\left.+y_1, y_2 y_3+y_1, \cdots, y_2y_k+y_1\right)J\,dy_1\,dy_2\right\} \\ &=\frac{\int_0^\infty\int_{-\infty}^\infty (x_1-\theta)p\left(\frac{x-\theta\mathbb{1}}{\sigma}\right)\sigma^{-k-3}\,d\theta\,d\sigma}{\int_0^\infty\int_{-\infty}^\infty \tilde{s}p\left(\frac{x-\theta\mathbb{1}}{\sigma}\right)\sigma^{-k-3}\,d\theta\,d\sigma} \end{aligned}$$

最后的等式是由分子分母的积分作变数变换，$(y_1, y_2)\to(\theta, \sigma)$：$y_1=(x_1-\theta)/\sigma$, $y_2=\tilde{s}/\sigma$，从而可知有

$$(x_2-\theta)/\sigma=y_2\left(1-\sum_{i=3}^k y_i^2\right)^{1/2}\operatorname{Sgn}(x_2-x_1).$$

而 $y_i(3\leqslant i\leqslant k)$ 以原来的 $(x_i-x_1)/\tilde{s}$ 代回，得 $y_1+y_2y_i=(x_i-\theta)/\sigma$，即得上面的积分等式.

同理可得

$$\frac{E_{1,0}(\tilde{s}|T)}{E_{1,0}(\tilde{s}^2|T)}=\frac{\int_0^\infty\int_{-\infty}^\infty \sigma^{-k-2}p\left(\frac{x-\theta 1}{\sigma}\right)d\theta\,d\sigma}{\tilde{s}\int_0^\infty\int_{-\infty}^\infty \sigma^{-k-3}p\left(\frac{x-\theta 1}{\sigma}\right)d\theta\,d\sigma}.$$

由(2.10)与(2.11)式即可得

$$\hat{\theta}=\frac{\int_0^\infty\int_{-\infty}^\infty \theta\sigma^{-k-3}p\left(\frac{x-\theta 1}{\sigma}\right)d\theta\,d\sigma}{\int_0^\infty\int_{-\infty}^\infty \sigma^{-k-3}p\left(\frac{x-\theta 1}{\sigma}\right)d\theta\,d\sigma}; \tag{2.12}$$

$$\hat{\sigma}=\frac{\int_0^\infty\int_{-\infty}^\infty \sigma^{-k-2}p\left(\frac{x-\theta 1}{\sigma}\right)d\theta\,d\sigma}{\int_0^\infty\int_{-\infty}^\infty \sigma^{-k-3}p\left(\frac{x-\theta 1}{\sigma}\right)d\theta\,d\sigma}. \tag{2.13}$$

例 2.5 问题如例 2.1. 参数空间为 $-\infty<\alpha<\infty$, $\beta>0$, 要估计 α、β, 损失函数为 $a(d_1-\alpha)^2/\beta^2+b(d_2-\beta)^2/\beta^2$. 根据(2.12)与(2.13)，基于 $x_{(1)}\leqslant x_{(2)}\leqslant\cdots\leqslant x_{(r)}$ 的 α、β 的最优同变估计为

$$\hat{\alpha}=\frac{\int_0^\infty\int_{-\infty}^{x_{(1)}}\beta^{-r-3}\alpha\exp\left(-\frac{U-n\alpha}{\beta}\right)d\alpha\,d\beta}{\int_0^\infty\int_{-\infty}^{x_{(1)}}\beta^{-r-3}\exp\left(-\frac{U-n\alpha}{\beta}\right)d\alpha\,d\beta}$$

$$=\frac{\int_0^\infty\beta^{-r-3}\left(\frac{\beta}{n}x_{(1)}-\frac{\beta^2}{n^2}\right)\exp\left(-\frac{U-nx_{(1)}}{\beta}\right)d\beta}{\int_0^\infty\frac{1}{n}\beta^{-r-2}\exp\left(-\frac{U-nx_{(1)}}{\beta}\right)d\beta}$$

$$=x_{(1)}-(U-nx_{(1)})/nr=x_{(1)}-\left(\sum_{i=1}^r x_{(i)}-rx_{(1)}\right)\Big/nr; \tag{2.14}$$

$$\hat{\beta}=\frac{\int_0^\infty\int_{-\infty}^{x_{(1)}}\beta^{-r-2}\exp\left(-\frac{U-n\alpha}{\beta}\right)d\alpha\,d\beta}{\int_0^\infty\int_{-\infty}^{x_{(1)}}\beta^{-r-3}\exp\left(-\frac{U-n\alpha}{\beta}\right)d\alpha\,d\beta}$$

$$=\frac{\int_0^\infty \beta^{-r-1}\exp\left(-\frac{U-nx_{(1)}}{\beta}\right)d\beta}{\int_0^\infty \beta^{-r-2}\exp\left(-\frac{U-nx_{(1)}}{\beta}\right)d\beta}$$

$$=(U-nx_{(1)})/r=\left(\sum_{i=1}^{r}x_{(i)}-rx_{(1)}\right)\Big/r. \tag{2.15}$$

将上述结果与α已知或β已知时的估计作比较，注意到$\alpha(U-nx_{(1)})/\beta$服从$\chi^2(r-1)$分布，且与$x_{(1)}$独立(见第二章例3.3)，易算得二估计的风险函数是

$$a(r+1)/n^2r+b/r. \tag{2.16}$$

用公式(2.12)与(2.13)还容易求得

i) 正态分布$N(\mu, \sigma^2)$，对$x_1, \cdots, x_n$为简单独立抽样结果，(μ, σ)的最优同变估计为

$$\left(\bar{x},\ \frac{\Gamma(n/2)\sqrt{2}}{\Gamma((n-1)/2)(n-1)}S_n\right).$$

这里$\bar{x}=\sum_{i=1}^{n}x_i/n$, $S_2^n=\sum_{i=1}^{n}(x_i-\bar{x})^2$，估计的风险函数为

$$\frac{b}{n}+a\left(1-\frac{2\Gamma^2(n/2)}{\Gamma^2((n-1)/2)}\frac{1}{(n-1)}\right).$$

ii) 当$(x_1, \cdots, x_n)$有均匀分布密度$\prod_{i=1}^{n}\left\{\frac{1}{\sigma}I_{\left(\theta-\frac{\sigma}{2},\theta+\frac{\sigma}{2}\right)}(x_i)\right\}$ $(n\geqslant 2)$时，(θ, σ)的最优同变估计是

$$(x_{(1)}+x_{(n)})/2,\ (x_{(n)}-x_{(1)})(n+2)/n.$$

这里$x_{(1)}\leqslant\cdots\leqslant x_{(n)}$是顺序统计量，估计风险函数为

$$\frac{b}{2(n+1)(n+2)}+\frac{2a}{(n+1)(n+2)}.$$

对于极值分布、Logistic分布、Cauchy分布、双指数分布等位置刻度分布族的一些分布，位置参数及刻度参数皆未知时的最优同变估计皆可用(2.12)与(2.13)来进行计算得出.

四、多维正态分布$N(\boldsymbol{\mu}, \boldsymbol{A})$, $\boldsymbol{\mu}\in\boldsymbol{R}^k$, $\boldsymbol{A}>0$, $\boldsymbol{\mu}$及$\boldsymbol{A}$的最优同变估计

设$x_1, \cdots, x_n (n\geqslant k)$是从总体的独立抽样的结果.

i) 关于μ的最优同变估计

采用损失函数$(\mu-d)^{\tau}A^{-1}(\mu-d)$，在例2.3给出当$A>0$已知时，采用转移群时，$\bar{x}$是$\mu$的最优同变估计．这里我们采用多维仿射变换群$\mathscr{G}$：

$$\mathscr{G}=\left\{\begin{pmatrix}1 & 0\\ e & C\end{pmatrix}: e\in R^k,\ \begin{array}{l}C\text{为}k\times k\text{维下三角阵,}\\ \text{且对角线上的元素为正}\end{array}\right\}$$

它在矩阵运算下构成一个群，对任意的$g\in\mathscr{G}$，

$$gx=(cx_1+e,\ cx_2+e,\ \cdots,\ cx_n+e)$$

由样本空间引导出参数空间的变换群$\overline{\mathscr{G}}=g$，$\forall\bar{g}\in\overline{\mathscr{G}}$，

$$\bar{g}(\mu,\ A)=(c\mu+e,\ cAc^{\tau}).$$

若行动空间采用与$\mathscr{G}$同构的变换群$\mathscr{G}^*=\mathscr{G}$，对于$\forall g^*\in\mathscr{G}^*$，$g^*d=cd+e$，则使此统计判决问题不变.

对于关于仿射群的μ的任一同变估计$\tilde{\mu}$，$\tilde{\mu}(gx)=g^*\tilde{\mu}(x)$，对$\forall g\in\mathscr{G}$成立．当取$g=\begin{pmatrix}1 & 0\\ \bar{x} & S\end{pmatrix}^{-1}$，此处$SS^{\tau}=\sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x})^{\tau}$，$S$是下三角阵且对角元素为正，且记$T=(S^{-1}(x_1-\bar{x}),\ \cdots,\ S^{-1}(x_n-\bar{x}))$．则

$$\begin{aligned}\tilde{\mu}(x)&=g^{*-1}\tilde{\mu}(gx)=\bar{x}+S\tilde{\mu}(S^{-1}(x_1-\bar{x}),\ \cdots,\ S^{-1}(x_n-\bar{x}))\\ &=\bar{x}+S\tilde{\mu}(T).\end{aligned}$$

此即μ的同变估计的一般形式．根据正态分布性质$\bar{x}$与T独立，从而也与S独立，故

$$\begin{aligned}&E\{(\tilde{\mu}(x)-\mu)^{\tau}A^{-1}(\tilde{\mu}(x)-\mu)\,|\,\mu,\ A\}\\ &=E\{(\bar{x}-\mu)^{\tau}A^{-1}(\bar{x}-\mu)\,|\,\mu,\ A\}\\ &\quad+E\{(S\tilde{\mu}(T))^{\tau}A^{-1}S\tilde{\mu}(T)\,|\,\mu,\ A\}\\ &\geqslant E\{(\bar{x}-\mu)A^{-1}(\bar{x}-\mu)\}=k/n.\end{aligned}$$

因此，$\bar{x}$是μ的最优同变估计.

ii) 关于协方差阵的最优同变估计

损失函数采用$\operatorname{tr}(A^{-1}D-I_k)^2(D>0)$，变换群采用在i)中所述的类似仿射群，但要求$C$为非退化的，因此$gx$仍为正态，只是参数变为$\bar{g}(\mu,\ A)$．又

$$L(\tilde{g}A,\ g^*D)=\operatorname{tr}((CAC^\tau)^{-1}CDC^\tau-I)^2$$
$$=\operatorname{tr}(C^{\tau-1}A^{-1}DC^\tau-I)^2=\operatorname{tr}(A^{-1}D-I)^2=L(A,\ D).$$

故这个统计估计是不变的. 注意，这损失函数为凸函数: 因为对 $A>0$ 的任意估计 $D_1>0,\ D_2>0$ 及 $0<p<1,\ q=1-p$, 有

$$\operatorname{tr}[A^{-1}(pD_1+qD_2)-I]^2$$
$$=\operatorname{tr}[p^2A^{-1}D_1A^{-1}D_1+q^2A^{-1}D_2A^{-1}D_2-2pA^{-1}D_1$$
$$-2qA^{-1}D_2+2pqA^{-1}D_1A^{-1}D_2+I]$$
$$=\operatorname{tr}[p(A^{-1}D_1-I)^2+q(A^{-1}D_2-I)^2-pq(A^{-1}(D_1-D_2))^2]$$
$$\leqslant\operatorname{tr}[p(A^{-1}D_1-I)^2+q(A^{-1}D_2-I)^2].$$

(注意 A^{-1} 可表为 $B'B$).

其次, 由于 $x=(x_1,\ \cdots,\ x_n)$ 的联合密度是

$$(2\pi)^{-kn/2}(\det A)^{-n/2}\exp\left[-\sum_{i=1}^n(x_i-\mu)^\tau A^{-1}(x_i-\mu)\Big/2\right]$$
$$=(2\pi)^{-kn/2}(\det A)^{-n/2}\exp[-n(\bar{x}-\mu)^\tau A^{-1}(\bar{x}-\mu)/2$$
$$-\operatorname{tr}A^{-1}S^2/2],$$

此处 $$\bar{x}=\sum_{i=1}^n x_i/n,\ S^2=\sum_{i=1}^n(x_i-\bar{x})(x_i-\bar{x})^\tau.$$

由此不难推得 $x_2-\bar{x},\ \cdots,\ x_n-\bar{x}$ 的密度函数是 $\text{const}\cdot\exp\times[-\operatorname{tr}A^{-1}S^2/2]$. 这就说明了 S^2 是 $x-\bar{x}\mathbf{1}$ 的充分统计量, 而 A 的同变估计满足 $D(x)=D(x-\bar{x}\mathbf{1}^\tau)$, 所以由充分统计量的优良性质(见第一章定理 **6.3**), 最优同变估计应在依赖于充分统计量 $S^2=SS^\tau$, $\bar{x}$ 的同变估计中寻找. 对此类同变估计, 取 $g=\begin{pmatrix}1&0\\ \bar{x}&S\end{pmatrix}^{-1}$, 由定义可得

$$D(\bar{x},\ S^2)=SD(0,\ I)S^\tau.$$

我们还可断言 $D(0,\ I)$ 仅为对角形. 因为若再取 $g=\begin{pmatrix}1&0\\0&\tilde{C}\end{pmatrix}$, $\tilde{C}$ 为 $k\times k$ 对角阵, 对角元素取值 $+1$ 或 -1, 对所有这类 $\mathscr{G}$ 应有下式成立:

$$D(0,\ I)=D(g(0,\ I))=\tilde{C}D(0,\ I)\tilde{C}^\tau.$$

若 $\tilde{C}$ 的第 i 个对角线元素与第 j 个对角线元素异号, 则 $\tilde{C}D(0,$

$I)\tilde{C}^{\tau}$ 的 (i, j) 及 (j, i) 元素与 $D(0, I)$ 相应元素异号，欲使 $D(0, I)=\tilde{C}D(0, I)\tilde{C}^{\tau}$，必须 $D(0, I)$ 的 (i, j) 及 (j, i) 元素为 0. 令 $D(0,I)$ 的对角元素为 $h_1, \cdots, h_k$，现考查同变估计 $SD(0, I)S^{\tau}$ 的风险函数

$$
\begin{aligned}
R(\mu, A; D) &= E\{\operatorname{tr}(A^{-1}D(\bar{x}, S^2)-I)^2|\mu, A\} \\
&= E\{\operatorname{tr}(D(\bar{x}, S^2)-I)^2|0, I\} \\
&= E_{0,I}(\operatorname{tr} SD(0, I)S^{\tau}SD(0, I)S^{\tau}-2\operatorname{tr} SD(0, I)S^{\tau}+k).
\end{aligned}
$$

记 S 之诸元素为 $S_{ij}(S_{ij}=0, i>j)$. 令 $S^{\tau}S=M$. M 的诸元素为 m_{ij}. 于是

$$
m_{ii}=\sum_{j\geq i}S_{ij}^2, \quad m_{i\nu}=\sum_{j\geq\nu}S_{ij}S_{\nu j} \quad k\geqslant\nu>i.
$$

从多元分析知识知，S^2 服从 Wishart 分布，维数为 k，自由度为 $n-1$，其密度为

$$
\pi^{\frac{-k(k-1)}{4}}\left(\Gamma\left(\frac{n-1}{2}\right)\cdots\Gamma\left(\frac{n-k}{2}\right)\right)^{-1}(2^k\det A)^{\frac{-n+1}{2}}
$$

$$
\times(\det S^2)^{\frac{n-k-2}{2}}\exp\left(-\frac{1}{2}\operatorname{tr} A^{-1}S^2\right)(\text{当 } S>0 \text{ 时}).
$$

由于 $S^2\to S$ 的 Jacobian 为 $2^kS_{11}^kS_{22}^{k-1}\cdots S_{kk}$，故 S 的概率分布密度当 $A=I$ 时，为

$$
\frac{S_{11}^{n-2}S_{22}^{n-3}\cdots S_{kk}^{n-k-1}\exp\left[-\frac{1}{2}\sum_{i=1}^{k}\sum_{j\geq i}^{k}S_{ji}^2\right]}{2^{\frac{(k-1)(n-1)}{2}}\pi^{\frac{k(k-1)}{4}}\cdot\Gamma\left(\frac{n-1}{2}\right)\cdots\Gamma\left(\frac{n-k}{2}\right)}
$$

$$
(\text{当 } S_{11}>0, \cdots, S_{kk}>0, -\infty<S_{ij}<\infty, 1\leqslant j<i\leqslant k).
$$

这就说明了对 $1\leqslant i<j\leqslant k$, $S_{ji}\sim N(0, 1)$ 及 $S_{ii}^2\sim\chi_{n-i}^2(1\leqslant i\leqslant k)$ 诸 S_{ij} 彼此独立. 这就是被称为 Bartlett 直角坐标分解的结果. 从而知 $m_{ii}\sim\chi_{n+k-2i}^2$. 所以

$$
Em_{ii}=ES_{ii}^2+E\sum_{i<j}S_{ij}^2=n-i+k-i=n+k-2i,
$$

$$
Em_{ii}^2=(n+k-2i)(n+k-2i+2) \quad (i=1, 2, \cdots, k),
$$

$$
Em_{ij}=E\left(\sum_{\nu>j}S_{i\nu}S_{j\nu}\right)=0 \quad (1\leqslant i<j\leqslant k),
$$

$$Em_{ij}^2 = ES_{jj}^2 S_{ij}^2 + \sum_{\nu>j} ES_{j\nu}^2 S_{i\nu}^2 + \sum_{\nu\neq l} ES_{j\nu} S_{i\nu} S_{jl} S_{il}$$
$$= n-j+k-j = n+k-2j \quad (1\leqslant i<j\leqslant k).$$

故当 $A=I$ 时，

$$E\,\mathrm{tr}\,SD(0,\ I)S^\tau = E\,\mathrm{tr}\,D(0,\ I)M = \sum_{i=1}^k Eh_i m_{ii}$$
$$= \sum_{i=1}^k h_i(n+k-2i).$$

$$\begin{aligned} &E\,\mathrm{tr}\,SD(0,\ I)S^\tau SD(0,\ I)S^\tau \\ &= E\,\mathrm{tr}\,D(0,\ I)MD(0,\ I)M = \sum_{i,j} Eh_i h_j m_{ij}^2 \\ &= \sum_{i=1}^k h_i^2(n+k-2i)(n+2k-2i+2) + 2\sum_{j>i} h_i h_j(n+k-2j) \\ &\triangleq h^\tau Bh, \end{aligned}$$

这里 $h=(h_1,\ \cdots,\ h_k)^\tau$; B 是对称矩阵

$$b_{ii}=(n+k-2i)(n+k-2i+2),\quad b_{ij}=n+k-2j,\quad 当\ j>i\ 时.$$

因为 $E\,\mathrm{tr}\,SD(0,\ I)S^\tau SD(0,\ I)S^\tau = h^\tau Bh>0$，当 $h\neq 0$ 时，故 $B>0$. 现在观察

$$R(\mu,\ A,\ D) = h^\tau Bh - 2\sum_{j=1}^k h_j(n+k-2j)+k.$$

要使 $R(\mu,\ A,\ D)$ 达到最小，当且仅当 h 满足方程

$$Bh=(n+k-2,\ n+k-4,\ \cdots,\ n-k)^\tau,$$
$$h=B^{-1}(n+k-2,\ n+k-4,\ \cdots,\ n-k)^\tau.$$

把 h 解出，代入 $SD(0,\ I)S^\tau$，即得 A 的最优同变估计.

当 $k=2$，$n\geqslant k$ 时，可以算得

$$h_{12}^{(n)} = \frac{n^2-(n-2)}{n^2(n+1)-(n-2)}>0,\quad h_{22}^{(n)} = \frac{n(n+1)}{n^2(n+1)-(n-2)}>0.$$

从而可以算出 h 的递推公式，并用归纳法证得 h 的各元素为正，且小于 1(证明留给读者作练习).

§3 位置参数最优同变估计的 minimax 性

在前一节，我们已探讨了在平方损失下位置参数最优同变估

计的求法和具体表达式. 在本节,我们研究在转移分布族下、在一般损失函数下,最优同变估计的存在性及 minimax 性. 50 年代时 Blackwell[2]、Girshick 及 Savage[6] 作过研究, 这里采用陈希孺[5]的这方面工作,对证明过程作了改进.

设 $X=(X_1, \cdots, X_n)$ 为随机矩阵, 其中 $X_i(i=1, 2, \cdots, n)$ 本身为 k 维随机向量 $x_i=(x_{i1}, \cdots, x_{ik})^\tau$, 记 $\theta=(\theta_1, \cdots, \theta_k)^\tau$, $\mathbb{1}=(1, \cdots, 1)^\tau_{1\times n}$, $\bar{x}=\sum_{i=1}^n x_i \Big/ n$, 假设 $X\sim P(x-\theta\mathbb{1}^\tau)$, P 的形式已知,$\theta\in R^k$. 我们的问题是估计 θ,其损失函数 $L(\theta, d)=W(\theta-d)$ 在样本空间的平移(或转移)群 $\mathscr{G}=\{g:g\in R^k\}$ 满足加法运算, 对 $\forall g\in\mathscr{G}$:

$$gx=x+g\mathbb{1}^\tau=(x_1+g, \cdots, x_n+g);$$

由它诱导出参数空间的平移变换群 $\bar{\mathscr{G}}=\mathscr{G}$, 对 $\forall g\in\mathscr{G}$, $\bar{g}\theta=\theta+g$; 同时,若行动空间采用平移变换群 $\mathscr{G}^*=\mathscr{G}$, 对 $\forall g^*\in\mathscr{G}^*$, $\mathscr{G}^*d=d+g$. 显然,这三个群是同构的. 又 $L(\bar{g}\theta, g^*d)=L(\theta, d)$,因此,参数 θ 的估计是在平移群的不变判决问题. 对于任一同变估计 $\hat{\theta}_I(x)$, 它应满足

$$\hat{\theta}_I(gx)=g^*\hat{\theta}_I(x)=\hat{\theta}_I(x)+g, \ \forall g\in\mathscr{G}.$$

若取 $g=-\bar{x}$, 立即得到

$$\hat{\theta}_I(x)=\bar{x}+\hat{\theta}_I(x-\bar{x}\mathbb{1}^\tau)\triangleq\bar{x}+S_I(x). \tag{3.1}$$

此处 $S_I(x)$ 满足:

$$S_I(x)=S_I(x+g\mathbb{1}^\tau), \ 对 \ \forall g\in\mathscr{G}, x\in\mathscr{X}.$$

对任何其它估计 $\hat{\theta}(x)=\bar{x}+S(x)$, 风险函数为

$$\begin{aligned}R(\theta, \hat{\theta}(x))&=\int_{R^{kn}}W(\theta-\hat{\theta}(x))\alpha P(x-\theta\mathbb{1}^\tau)\\&=\int_{R^{kn}}W[-S(x+\theta\mathbb{1}^\tau)-\bar{x}]\alpha P(x)\\&\triangleq\rho(\theta, S, P, W).\end{aligned} \tag{3.2}$$

若 $\hat{\theta}(x)$ 是同变估计,则 $R(\theta, \hat{\theta}_I(x))=\rho(0, S, P, W)$, 它与 θ 无关.

设 $\mathfrak{M}$ 为一切定义在 R^{kn} 上取值在 R^k 内的 Borel 向量可测

函数，$\mathscr{I}$ 为 $\mathfrak{M}$ 中一切满足(3.1)的函数的族. 定义

$$V=\inf_{S\in\mathscr{I}}\rho(0,\ S,\ P,\ W);$$

$$\bar{V}(W,\ P)=\inf_{S\in\mathfrak{M}}\sup_{\theta}\rho(\theta,\ S,\ P,\ W).$$

令 $G(\theta)$ 为 θ 的先验分布，记

$$\underline{V}(W,\ P)=\sup_{G}\inf_{S\in\mathfrak{M}}\int\rho(\theta,\ S,\ P,\ W)dG(\theta).$$

显然，$V\geqslant\bar{V}(W,\ P)\geqslant\underline{V}(W,\ P)$. 我们将证明 $V=\bar{V}(W,\ P)=\underline{V}(W,\ P)$. 同时解决最优同变估计的存在性问题.

关于损失函数的基本假设是:

1.° W 在 R^k 上连续，$W(0)=0$. 若 $P(x)$ 有密度函数 $p(x)$，则仅假设 $W(\cdot)$ 在有界区域上有界.

2.° $\lim\limits_{\|\theta\|\to\infty}W(\theta)=\infty$，或者对任意的 θ，有 $W(\theta)<\lim\limits_{\|\theta\|\to\infty}W(\theta)$.

由 1°、2° 知存在 $r>0$，使 $W(\theta)>0$(当 $\|\theta\|>r$).

3.° $W(t\theta)$ 作为实变量 t 的函数，在 $t\geqslant0$ 时非降，$t<0$ 时非增.

定理3.1 若损失函数满足 1°～3°，则存在最优同变估计 $\hat{\theta}_{I_0}(x)$，而且是 minimax 估计，即

$$R(\theta,\ \hat{\theta}_{I_0}(x))=V=\bar{V}(W,\ P)=\underline{V}(W,\ P).$$

在证明定理之前，先证明几条引理:

引理3.1 若 $W(t)$ 为 R^k 上的可测函数，在有界区域上有界，满足假设 2°、3° 及 $W(0)=0$，并存在 $m>0$，使得 $P(\|\bar{X}\|>m|\theta=0)=0$，则

$$V=\inf_{S\in\mathscr{I}}\rho(\theta,\ S,\ P,\ W)=\underline{V}.$$

证明 由假设 2°，$W(\theta)\leqslant\lim\limits_{\|\theta\|\to\infty}W(\theta)$，故存在充分大的 $a>3$，使得

$$\text{当 }\|\theta_1\|>(a-1)m,\ \|\theta_2\|\leqslant2m\text{ 时}, W(\theta_1)\geqslant W(\theta_2). \tag{3.3}$$

由于 $P(\|\bar{X}\|>m|\theta=0)=0$ 及分布函数属于位置参数分布族，故 $P(\|\bar{X}-\theta\|\leqslant m|\theta)=1$，由此可见，从抽样结果计算出 $\bar{x}$，θ 以概

率为1在以 $\bar{x}$ 为中心、m 为半径的球内，θ 的一个合理的估计 $\bar{x}+S(x)$，其值域应在此球内，也即满足关系式

$$\|\bar{x}-\bar{x}-S(x)\|=\|S(x)\|\leqslant m.$$

从这个直观分析出发，指导我们作如下设想：对于 θ 的任一估计 $\bar{x}+S(x)$，我们定义

$$S^*(x)=\begin{cases} S(x), & \text{当 } \|S(x)\|\leqslant am; \\ mS(x)/\|S(x)\|, & \text{当 } \|S(x)\|>am. \end{cases}$$

可以证明 θ 的这个估计 $\bar{x}+S^*(x)$ 优于 $\bar{x}+S(x)$. 这是因为当 $\|S(x+\theta\mathbf{1}^{\tau})\|\geqslant am$ 时，$\|S^*(x+\theta\mathbf{1}^{\tau})\|\leqslant m$. 又 $P_0(\|\overline{X}\|\leqslant m)=1$，故

$$\|\bar{x}+S(x+\theta\mathbf{1}^{\tau})\|\geqslant\|S(x+\theta\mathbf{1}^{\tau})\|-\|\bar{x}\|\geqslant(a-1)m, \text{ a. s. } P_0,$$

$$\|x+S^*(x+\theta\mathbf{1}^{\tau})\|\leqslant\|S^*(x+\theta\mathbf{1}^{\tau})\|+\|\bar{x}\|\leqslant 2m, \text{ a. s. } P_0.$$

再由(3.3)式，得

$$W(-\bar{x}-S(x+\theta\mathbf{1}^{\tau}))\geqslant W(-\bar{x}-S^*(x+\theta\mathbf{1}^{\tau})). \text{ a. s. } P_0.$$

在 $\|S(x+\theta\mathbf{1}^{\tau})\|\leqslant am$ 时，上式也成立，故由(3.2)式即知：$\rho(\theta, S, P, W)\geqslant\rho(\theta, S^*, P, W)$. 因此，令 S 表示所有满足 $\|S^*(x)\|\leqslant am$ 的统计量，则可把估计类限制在 $\hat{\theta}(x)=\bar{x}+S^*(x)$，$S^*(x)\in S$，并不影响引理的结论.

下面我们采用 θ 的先验分布作为 $\|\theta\|\leqslant T$ 上的均匀分布，由 $\underline{V}$ 的定义，对任给的 $\varepsilon>0$，存在 $S^*(x)\in S$ 或估计量 $\hat{\theta}^*(x)=\bar{x}+S^*(x)$，使得

$$\begin{aligned}\underline{V}+\varepsilon &\geqslant\frac{1}{V_T}\int_{\|\theta\|<T}d\theta\int W(-S^*(x+\theta\mathbf{1}^{\tau})-\bar{x})dP(x)\\ &=\frac{1}{V_T}\int dP(x)\int_{\|\theta\|<T}W(-S^*(x+\theta\mathbf{1}^{\tau})-\bar{x})d\theta.\end{aligned} \tag{3.4}$$

此处 V_T 表示 $\{\theta:\|\theta\|\leqslant T\}$ 的体积，其值等于 $\pi^{k/2}T^k/\Gamma((k+2)/2)$. 我们记

$$A=\{\theta:\|\theta\|\leqslant T, \|\theta+\bar{x}\|>T\};$$

$$B=\{\theta:\|\theta\|>T, \|\theta+\bar{x}\|\leqslant T\}.$$

因为 $\|S^*\|\leqslant am$，$\|\bar{x}\|\leqslant m$，a. s. P_0，故

$$\left|\frac{1}{V_T}\left(\int_A-\int_B\right)W(-S^*(x+\theta 1^\tau)-\bar{x})d\theta\right|$$
$$\leqslant\frac{1}{V_T}\left(\max_{|t|<(a+1)m}W(t)\right)\left(\int_A d\theta+\int_B d\theta\right)$$
$$\leqslant\frac{2M}{V_T}\int_{\{|\theta|<T+m,\ |\bar{x}+\theta|>T\}}d\theta$$
$$=2M[(T+m)^k-T^k]/T^k=2M[(1+T^{-1}m)^k-1].$$

此处 $M=\max\limits_{|t|<(a+1)m}W(t)$，因此继(3.4)式，有

$$\frac{1}{V_T}\int dP(x)\int_{|\theta|<T}W(-S^*(x+\theta 1^\tau)-\bar{x})d\theta$$
$$=\frac{1}{V_T}\int dP(x)\left(\int_{|\theta+\bar{x}|<T}+\int_A-\int_B\right)W(-S^*(x+\theta 1^\tau)-\bar{x})d\theta$$
$$\geqslant\frac{1}{V_T}\int dP(x)\int_{|\theta+\bar{x}|<T}W(-S^*(x+\theta 1^\tau)-\bar{x})d\theta$$
$$-2M[(1+T^{-1}m)^k-1]$$
$$=\frac{1}{V_T}\int_{|t|<T}dt\int W(-S^*(x+(t-\bar{x})1^\tau)-\bar{x})dP(x)$$
$$-2M[(1+T^{-1}m)^k-1]$$
$$\geqslant\inf_{|t|<T}\int W(-S^*(x+(t-\bar{x})1^\tau)-\bar{x})dP(x)$$
$$-2M[(1+T^{-1}m)^k-1]$$
$$\geqslant\inf_{S\in\mathscr{I}}\int W(-S^*(x)-\bar{x})dP(x)$$
$$-2M[(1+T^{-1}m)^k-1].\qquad(3.5)$$

上式第二个等式是作变换 $\theta+\bar{x}=t$，然后交换积分限．最后一个不等式是因为对固定的 t，$S^*(x+(t-\bar{x})1^\tau)$ 是一个不变统计量，而 $\mathscr{I}$ 表示所有不变统计量的集合类．根据(3.4)与(3.5)式，再令 $T\to\infty$，则立即得到 $\underline{V}+\varepsilon\geqslant V$．由于 ε 的任意性，故 $\underline{V}\geqslant V$．前面已指出有 $\underline{V}\leqslant\bar{V}\leqslant V$，因此 $\underline{V}=V$，引理得证．

注 1　对于 P 是概率测度的假定若放宽为有限测度，结论仍然成立．

注 2　若 $W(\theta)$ 有界，则引理结论仍然成立．

引理3.2 若假设1°、2°成立，μ为$(R^k, \mathscr{B}_k)$上的σ-有限测度，而且对任意有界集B，有$\mu(B)<\infty$。令$h(a)=\int W(a+\theta)d\mu$，又若

$$h_m(a) \triangleq \int_{\|\theta\|<m} W(a+\theta)d\mu$$

对所有a连续（对一切m），则

$$\lim_{m\to\infty}\inf_a h_m(a)=\inf_a h(a). \tag{3.6}$$

当上式右端有限时，则存在有限a_0，使右端的下确界在a_0达到，即$h(a_0)=\inf h(a)$。

证明 我们取m充分大，致$\mu(\|\theta\|\leqslant m)>0$，由于$W(\theta)<\lim\limits_{\|\theta\|\to\infty} W(\theta)$及假设1°，故$h_m(a)$最小值必在某有限点$a_m$达到，由定义$h_m(a_m)\geqslant b_{m-1}(a_m)\geqslant b_{m-1}(a_{m-1})$，故$h_m(a_m)$是$m$的非降函数。显然，

$$A\triangleq \lim_{m\to\infty} h_m(a_m)\leqslant \inf_a \int W(a+\theta)d\mu. \tag{3.7}$$

下面证上式等号成立：若$A=\infty$，则引理正确，若$A<\infty$，这时$\{a_m\}$必有界．否则，就存在子序列$\{a_{m_k}\}$，使$\|a_{m_k}\|\to\infty(k\to\infty)$。这时对任何的$b<\infty$，由Fatou引理知，对任意的有限$b$，有

$$A=\lim_{k\to\infty} h_{m_k}(a_{m_k})\geqslant \lim_{k\to\infty}\inf \int_{\|\theta\|<b} W(a_{m_k}+\theta)d\mu$$

$$\geqslant \int_{\|\theta\|<b} \lim_{k\to\infty}\inf W(a_{m_k}+\theta)d\mu \geqslant \int_{\|\theta\|<b} \lim_{\|\theta\|\to\infty} W(\theta)d\mu.$$

令$b\to\infty$，再注意假设2°，由上式可得

$$\text{对于}\ \forall a,\quad A\geqslant \int \lim_{\|\theta\|\to\infty} W(\theta)d\mu > \int W(\theta+a)d\mu.$$

因而$A>h(a)\geqslant \inf h(a)$，这与(3.7)式矛盾，说明$\{a_m\}$有界，从而存在子序列$\{a_{m_k}\}$，使$a_{m_k}\to a^\circ(k\to\infty)$．注意当$m_k$充分大时，$m_k\geqslant b$（可取$b$为整数），

$$h_b(a_{m_k})=\int_{\|\theta\|<b} W(a_{m_k}+\theta)d\mu\leqslant \int_{\|\theta\|<m_k} W(a_{m_k}+\theta)d\mu$$

$$=h_{m_k}(a_{m_k})\leqslant A<\infty.$$

因 $h_m(a)$ 是 a 的连续函数，即有 $h_b(a^0)\leqslant A$，从而 $h(a^0)\leqslant A<\infty$，与(3.7)比较，有 $A=h(a^0)=\inf h(a)$．引理证完．

注 若 $\lim\limits_{\|\theta\|\to\infty} W(\theta)$ 不存在，则引理不一定成立，例如取 $W(\theta)=\theta I_{(\theta\geqslant 0)}$，$d\mu(x)=[\pi(1+\theta)^2]^{-1}d\theta$．因为 $\int_{-m}^{m} W(-m+\theta)d\mu(\theta)=0$．故引理不成立．

引理 3.3 假设

i) μ 为 $(R^k, \mathscr{B}_k)$ 上的测度，对任意有界集 B，$\mu(B)<\infty$，且有密度函数 $d\mu=g(\theta)d\theta$．

ii) $W(\theta)$ 在任意有界区域上有界可测，

则
$$h_m(a)=\int_{|\theta|<m} W(a+\theta)g(\theta)d\theta$$
是 a 的连续函数．

证明 任取一串 $a_n\to a(n\to\infty)$，不失一般性，令 $\|a_n-a\|<\delta$ $(n=1, 2, \cdots)$ 在 $\|\theta\|\leqslant m+\delta$ 上，根据 Ruzin 定理(参见 И. П. Натансон《实变函数论》第四章定理 4)，对任意的 $\varepsilon>0$，存在连续函数 $f_{a\varepsilon}(\theta)$，$|f_{a\varepsilon}(\theta)|\leqslant \max\limits_{|\theta|<m+\delta} W(a+\theta)$，使

$$L\{\theta: f_{a\varepsilon}(\theta)\neq W(a+\theta),\ \|\theta\|\leqslant m+\delta\}\triangleq L\{B_{a\varepsilon}\}\leqslant \varepsilon/2.$$

故有

$$\begin{aligned}\|h_m(a)-h_m(a_n)\|&\leqslant\int_{|\theta|<m}|W(a+\theta)-W(a_n+\theta)|g(\theta)d\theta\\&\leqslant 2\int_{B_{a\varepsilon}} g(\theta)d\theta \sup_{|\theta|<m+\delta} W(a+\theta)\\&\quad+\int_{|\theta|<m}|f_{a\varepsilon}(\theta)-f_{a\varepsilon}(a_n-a+\theta)|g(\theta)d\theta.\end{aligned}$$

由于 $f_{a\varepsilon}(\theta)$ 在 $\|\theta\|\leqslant m+\delta$ 上一致连续及 $\int_{\|\theta\|<m} g(\theta)d\theta<\infty$，故

$$\lim_{n\to\infty}|h_m(a)-h_m(a_n)|\leqslant 2\sup_{\|\theta\|<m+\delta} W(a+\theta)\int_{B_{a\varepsilon}} g(\theta)d\theta.$$

由于 ε 的任意性，上式右端任意小，引理得证．

引理 3.4 令 $\mathscr{B}_k$ 为 $\mathscr{R}^k$ 上的所有 Borel 集合类，$D\in\mathscr{B}_k$．设 $P(y, A)$ 为定义在 $D\times\mathscr{B}_k$ 上的函数，满足下列条件：

1° 对固定的 $y\in D$, 它是 $\mathscr{B}_k$ 上的概率测度;

2° 对固定的 $A\in\mathscr{B}_k$, 它是 D 上 Borel 可测函数, 又设 $W(\theta)$ 满足引理 3.2 的条件, 则对任意的正整数 m, 存在定义在 D 上 Borel 可测的函数 $a_m(y)$, 使

$$\int_{|x|<m} W(-a_m(y)-x)P(y,\ dx)$$
$$=\inf_a\int_{|x|<m} W(-a-x)P(y,\ dx). \qquad (3.8)$$

又存在 D 上 Borel 可测函数 $a(y)$, 使

$$\int W(-a(y)-x)P(y,\ dx)=\inf_a\int W(-a-x)P(y,\ dx). \quad (3.9)$$

从引理 3.2 的结论及证明中可知存在函数 $a_m(y)$ 及 $a(y)$ 满足要求, 至于它们的可测性, 这里就不讨论了. 有兴趣的读者可参见[5].

定理的证明 很易验证, 引理 3.1～3.4 的条件是满足的. 取 $Z(x)=x-\bar{x}1^\tau$, 给定 $Z(x)$, $\bar{x}$ 的正则条件概率函数为 $P(z,\ d\bar{x})$. 这时对任何非负的 Borel 可测函数 $\psi(X)$, 有

$$E\{\psi(X)\,|\,Z\}=E\{\psi(Z+\bar{x}1^\tau)\,|\,Z\}$$
$$=\int\psi(Z+\bar{x}1^\tau)P(Z,\ d\bar{x}).$$

在此及以下, E 皆指 E_0.

若 $f\in\mathscr{I}$, 则 $f(x)=f(Z)$, 此时

$$E\{W(-f(x)-\bar{x})\,|\,Z\}=\int W(-f(z)-\bar{x})P(z,\ d\bar{x}),$$

根据引理 3.4, 取 $D=R^k$, 存在 Borel 可测的函数 $f_0(z)$, 使

$$\overline{W}(Z)=E\{W(-f_0(Z)-\bar{x})\,|\,Z\}=\inf_a E\{W(-a-\bar{x})\,|\,Z\},$$

因此对任何 $f\in\mathscr{I}$, 皆有

$$\rho(f)=E\{W(-f(Z)-\bar{x})\}$$
$$=E\{E[W(-f(Z)-\bar{x})\,|\,Z]\}\geqslant E\overline{W}(Z)=\rho(f_0).$$

故不论 $V=\infty$ 或 $V<\infty$, 皆有

$$V=V(P,\ W)=\inf_{f\in\mathscr{I}}\rho(f)=\rho(f_0). \qquad (3.10)$$

现在设m为自然数，定义测度P_m如下:

$$P_m(A)=P(A\cap\{x:\|\bar{x}\|\leqslant m\}),$$

又设$h(x, m)$为集合$\{x:\|\bar{x}\|\leqslant m\}$的示性函数，这时对任何$f\in\mathscr{I}$，有

$$E\{h(x, m)W(-f(Z)-\bar{x})\,|\,Z\}$$
$$=\int_{\|\bar{x}\|<m}W(-f(Z)-\bar{x})P(Z, d\bar{x}).$$

根据引理3.4，存在Borel可测的函数$f_m(Z)$，使

$$\overline{W}_m(Z)=E\{h(x, m)W(-f_m(Z)-\bar{x})\,|\,Z\}$$
$$=\inf_a E\{h(x, m)W(-a-\bar{x})\,|\,Z\}.$$

同导致(3.10)类似的推理可得

$$V_m=V(P_m, W)=\rho(f_m, P_m, W)=E\overline{W}_m(Z). \tag{3.11}$$

由引理3.2，知$\overline{W}_m(Z)\uparrow W(Z)$，$m\to\infty$. 根据单调收敛定理，有

$$\lim_{m\to\infty}V_m=\lim_{m\to\infty}E\overline{W}_m(Z)=EW(Z)=V.$$

由引理3.1知$V_m=\underline{V}(P_m, W)$，且显然

$$\underline{V}=\underline{V}(P, W)\geqslant\underline{V}(P_m, W).$$

综合上述结果，即得

$$\underline{V}\geqslant\underline{V}(P_m, W)=V_m\to V\,(m\to\infty),$$

即$\underline{V}\geqslant V$，再由于$V\geqslant\overline{V}\geqslant\underline{V}$，得$\underline{V}=V$. 由上述证明可看出：具有转移性质的位置参数的minimax估计可取为$\bar{x}+f_0(Z)$，定理证完.

注1 如果参数不是全空间R^k，而是其真子集，只要空间Θ满足:

$$\lim_{T\to\infty}\frac{T^{k-1}}{V(S_T\cap\Theta)}=0,$$

定理仍然成立. 这里$S_T=\{\theta:\|\theta\|<T\}$. 为证明这一点，只要将引理3.1的先验分布取成$S_T\cap\Theta$上的均匀分布即可.

注2 若$W(\theta)$是严格凸函数，且存在风险有限的估计，则同变的minimax估计是唯一的.

例3.1 令$X\sim N(\mu, A)$，$A>0$已知，而$\mu\in R^k$未知. 损失

函数为 $W(\mu-d)$, $W(\theta)$ 为凸函数, 而且是 θ 各变量的对称函数, $W(\theta)\to\infty(\|\theta\|\to\infty)$. 则 x 是 μ 的同变 minimax 估计. 为此, 根据定理 3.1, 只要证明 x 是平移群下的最优同变估计. 易知, 任一同变估计形为 $x+C$, $C\in R^k$, 其风险函数为

$$\varphi(C)=E_0W(x+C)=(2\pi)^{-k/2}(\det A)^{-1/2}\times\int W(y+C)\exp\left[-\frac{1}{2}y^\tau A^{-1}y\right]dy.$$

由 $W(\theta)$ 的凸性得 $\varphi(C)$ 是 C 的凸函数. 由 $W(\theta)$ 的对称性及 $\exp[-y^\tau A^{-1}y/2]$ 关于 y 对称, 得 $\varphi(C)$ 是 C 的对称函数, 因此

$$\varphi(0)\leqslant[\varphi(C)+\varphi(-C)]/2=\varphi(C).$$

这就说明了 x 是最优同变估计.

例 3.2 设 $X\sim N(\mu, A)$, $\mu\in R^k$, $A>0$ 皆未知, 要求估计 μ, 损失函数为 $W((\mu-d)^\tau A^{-1}(\mu-d))$, 对于 $t\geqslant0$ $W(t)$ 是 t 的非降函数, 且 $W(t)<\lim\limits_{t\to\infty}W(t)$, 则 x 是在平移群的最优同变估计, 也是 minimax 估计.

首先假设 A 已知, 此时同变估计为 $x+C$, 其风险函数为

$$\varphi(C)=E\{W((x+C-\mu)^\tau A^{-1}(x+C-\mu))|\mu, A\}.$$

令 P 为正交变换, 化 A 至主轴, 具体地令

$$PAP^\tau=\begin{pmatrix}a_{11} & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & a_{kk}\end{pmatrix},$$

$$Z=\begin{pmatrix}\sqrt{a_{11}} & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \sqrt{a_{kk}}\end{pmatrix}^{-1}P(x-\mu),$$

$$C'=\begin{pmatrix}\sqrt{a_{11}} & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \sqrt{a_{kk}}\end{pmatrix}PC,$$

则

$$\varphi(C)=(2\pi)^{-k/2}(\det(A))^{-1/2}$$
$$\times\int_{R^k}W[(x+C-\mu)^\tau A^{-1}(x+C-\mu)]$$
$$\times\exp[-(x-\mu)^\tau A^{-1}(x-\mu)/2]dx$$
$$=(2\pi)^{-k/2}\int_{R^k}W(\|Z+C'\|^2)\exp[-\|Z\|^2/2]dZ\triangleq\bar{\varphi}(C').$$

易知 $\bar{\varphi}(C')$ 为球对称，因为对任意的正交变换 Q，有

$$\bar{\varphi}(QC')=(2\pi)^{-k/2}\int_{R^k}W(\|Z+QC'\|^2)e^{-\|z\|^2/2}dZ$$
$$=(2\pi)^{-k/2}\int_{R^k}W(\|Q^\tau Z+C'\|^2)e^{-\|z\|^2/2}dZ$$
$$=(2\pi)^{-k/2}\int_{R^k}W(\|y+C'\|^2)e^{-\|y\|^2/2}dy=\bar{\varphi}(C').$$

因此 $\bar{\varphi}(C')$ 可表为 $\psi(\|C'\|)$

$$\psi(\|C'\|)-\psi(0)=(2\pi)^{-k/2}\int_{R^k}\Big[W\Big(\sum_{i=2}^k y_i^2+(y_1+\|C'\|)^2\Big)$$
$$-W(\|y\|^2)\Big]e^{-\|y\|^2/2}dy$$
$$=(2\pi)^{-k/2}\Big(\int_{y_1>-\|C'\|/2}+\int_{y_1<-\|C'\|/2}\Big)$$
$$\Big[W\Big(\sum_{i=2}^k y_i^2+(y_1+\|C'\|)^2\Big)-W(\|y\|^2)e^{-\|y\|^2/2}dy$$
$$=(2\pi)^{-k/2}\int_{z_1<\|C'\|/2}\Big[W\Big(\sum_{i=2}^k z_i^2+(z_1-\|C'\|)^2\Big)-W(\|z\|^2)\Big]$$
$$\times e^{-\|z\|^2/2}dz-(2\pi)^{-k/2}\int_{z_1<-\|C'\|/2}\Big[W(\|y\|^2)$$
$$-W\Big(\sum_{i=2}^k y_i^2+(y_1+\|C'\|)^2\Big)\Big]e^{-\|y\|^2/2}dy$$
$$=(2\pi)^{-k/2}\int_{y_1<\|C'\|/2}\Big[W\Big(\sum_{i=2}^k y_i^2+(y_1+\|C'\|)^2\Big)+W(\|y\|^2)\Big]$$
$$\times[e^{-\|y\|^2/2}-e^{-[\sum_{i=2}^k y_i^2+(y_1+\|C'\|)^2]/2}]dy\geqslant 0.$$

因为 $W(t)\uparrow$，而 $\exp[-t^2/2]\downarrow$（当 $t>0$），所以 $\varphi(C)>\varphi(0)$.
这就说明了 x 是最优同变估计，根据定理 3.1，它也是 minimax 估计．由于 x 与 $A>0$ 取何值无关，风险函数也与 A 无关，因此在

A 未知时，结论仍然成立.

例 3.3 设 $y=x\beta+\varepsilon$, $\varepsilon\sim N(0,\ \sigma^2 I_n)$ $(n\geqslant k)$, k 为 β 的维数，x 已知而且 $x^\tau x>0$, $\beta\in R^k$ 未知. 要估计 β, 损失函数为 $W((\beta-d)/\sigma)$.

假设 $W(t)$ 是凸函数，且 $W(-t)=W(t)$, 则最小二乘估计 $\hat{\beta}=(x^\tau x)^{-1}x^\tau y$ 是 β 的 minimax 估计.

证明
$$
\begin{aligned}
&(y-x\beta)^\tau(y-x\beta)\\
&=[y-x(x^\tau x)^{-1}x^\tau y+x(x^\tau x)^{-1}x^\tau y-x\beta]^\tau\\
&\quad\times[y-x(x^\tau x)^{-1}x^\tau y+x(x^\tau x)^{-1}x^\tau y-x\beta]\\
&=y^\tau[I-x(x^\tau x)^{-1}x^\tau]^\tau[I-x(x^\tau x)^{-1}x^\tau]y\\
&\quad+[(x^\tau x)^{-1}x^\tau y-\beta]^\tau x^\tau x[(x^\tau x)^{-1}x^\tau y-\beta].
\end{aligned}
$$
在已知 σ^2 的条件下，$(x^\tau x)^{-1}x^\tau y$ 是充分统计量以及 $W(t)$ 为凸函数，由例 3.1，在平移群，$(x^\tau x)^{-1}x^\tau y$ 是 β 的最优同变估计，也是 minimax 估计. 但
$$
\begin{aligned}
&E\{W(((x^\tau x)^{-1}x^\tau y-\beta)/\sigma)\,|\,\sigma,\ \beta\}\\
&=(2\pi)^{-k/2}\int_{R^k}W(y)\exp[-y^\tau x^\tau xy/2]\,dy.
\end{aligned}
$$
它与 σ 无关，因此在未知 σ 时，$\hat{\beta}=(x^\tau x)^{-1}x^\tau y$ 也是 minimax 估计.

§4 位置刻度参数最优同变估计的 minimax 性

在这一节将讨论与上节类似的问题. 只是分布函数和参数有变化. 证明的思路也相似.

设随机矩阵 $X=(X_1,\ \cdots,\ X_n)\sim P((x-\theta 1^\tau)/\sigma)$. 我们的问题是根据从 X 的抽样结果 x 来估计未知位置参数 $\theta\in R^k$ 和刻度参数 $\sigma>0$. 样本空间的变换群是仿射群 $\mathscr{G}=\{(a,\ b):a>0,\ b\in R^k\}$. 对于 $\forall g\in\mathscr{G}$, $gx=ax+b1^\tau$. 由此群诱导出参数空间 $((\sigma,\ \theta):\sigma>0,\ \theta\in R^k)$ 的变换群 $\overline{\mathscr{G}}=\mathscr{G}$. 对于 $\forall\bar{g}\in\overline{\mathscr{G}}$, $\bar{g}(\sigma,\ \theta)=(a\sigma,\ a\theta+b)$. 若估计的损失函数采用 $W(d_1/\sigma,\ (d_2-\theta)/\sigma)$,

则当行动空间也采用相应的仿射群 $\mathscr{G}^*=\mathscr{G}$ 时，我们的统计判决问题不变.

(σ, θ)的估计$(U(x), V(x))$被称为同变估计，如果对于任意的 $g\in\mathscr{G}$ 下式成立：

$$U(gx)=U(ax+b\mathbb{1}^\tau)=aU(x)=g^*U(x);$$
$$V(gx)=V(ax+b\mathbb{1}^\tau)=aV(x)+b=g^*V(x).$$

令(U_0, V_0) $(U_0>0)$是任意给定的同变估计，对于(σ, θ)的任意估计 $(U(x), V(x))$，我们定义

$$S(x)=U(x)/U_0(x),\ r(x)=(V(x)-V_0(x))/U_0(x). \quad (4.1)$$

那么(U, V)与(S, r)形成一一对应，且若$(U(x), V(x))$为同变估计，则

$$\begin{aligned} S(gx)&=S(ax+b\mathbb{1}^\tau)=S(x);\\ r(gx)&=r(ax+b\mathbb{1}^\tau)=r(x). \end{aligned} \quad (4.2)$$

对 $\forall g\in\mathscr{G}$ 成立，也即 $(S(x), r(x))$ 在上述仿射群变换下是不变的. 记 $\mathscr{I}$ 为所有这类统计量. 对任意由 $S(x)$、$r(x)$对应的 $U(x)$、$V(x)$，其风险函数为

$$\begin{aligned}\rho(S, r; \sigma, \theta) &\triangleq \int W\left(\frac{U(x)}{\sigma}, \frac{V(x)-\theta}{\sigma}\right)dP\left(\frac{x-\theta\mathbb{1}^\tau}{\sigma}\right)\\ &=\int_{R^{kn}} W[S(\sigma y+\theta\mathbb{1}^\tau)U_0(y),\\ &\qquad r(\sigma y+\theta\mathbb{1}^\tau)U_0(y)+V_0(y)]dP(y).\end{aligned}$$

令

$$V=V(P, W)=\inf_{(S,r)\in\mathscr{I}}\sup_{(\sigma,\theta)\in\Theta}\rho(S, r; \sigma, \theta; P, W);$$
$$\bar{V}=\bar{V}(P, W)=\inf_{(S,r)}\sup_{(\sigma,\theta)\in\Theta}\rho(S, r; \sigma, \theta; P, W);$$
$$\underline{V}=\underline{V}(P, W)=\sup_{G}\inf_{(S,r)}\int\rho(S, r; \sigma, \theta; P, W)dG.$$

此处 G 是(σ, θ) 的先验分布，Θ 是参数空间 $\{\sigma>0, \theta\in R^k\}$. 为了讨论简便，下面仅考虑 $k=1$ 的情形，不难把结论推广到一般 k 的情形.

对损失函数假设

1°. $W(t_1, t_2)$ 在 Θ 上连续，若 P 有概率密度，则仅要求

$W(t_1, t_2)$ 在 Θ 的任一闭区域上有界.

2°. $W(1, 0)=0$.

3°. 当 $t_1>0$ 确定时, $W(t_1, t_2)$在 $t_2\leqslant 0$ 非增,在 $t_2\geqslant 0$ 非降;在 t_2 固定时, $W(t_1, t_2)$ 在 $0\leqslant t_1\leqslant 1$ 非增,在 $t_1\geqslant 1$ 非降.

4_1°. $W(t_1, t_2)\to\infty$, 当 $t_1^{-1}+t_1+|t_2|\to\infty$ 一致成立;

4_2°. $W(t_1, t_2)\to\infty$, 当 $t_1+|t_2|\to\infty$(例如$(t_2-1)^2+t_2^2$);

4_3°. $W(t_1, t_2)$在 Θ 上连续, $W(t_1, t_2)\to a<\infty$, 当 $t_1+|t_2|\to\infty$ 一致成立.

记 $A=\{x:x_1=x_2=\cdots=x_n\}$, 显见 A 为不变集, 我们假设 $P(A)=0$, 对此,可选

$$U_0(x)=\max_{1\leqslant i<j\leqslant n}|x_i-x_j|,\ V_0(x)=x_1.$$

如果 $P(A)>0$, 易见在条件 4_1° 下, 任何估计的风险为∞, 若 $P(A)>0$ 且 P 为离散分布, 在条件 4_2° 下也可证明最优同变估计存在,且为 minimax 估计.

引理4.1 设 $P(A)=0$, 且存在 m, 使

$$P((U_0(x), V_0(x))\in E_m|1, 0)=1. \tag{4.3}$$

此处 $$E_m=\{(t_1, t_2):m^{-1}\leqslant t_1\leqslant m,\ |t_2|\leqslant m\}.$$

若损失函数满足假设 2°、3° 及 4° 之一, $W(t_1, t_2)$ 在有界区域上有界,则 $\underline{V}=V$.

证明 由(4.3)式,有

$$\begin{aligned}&P((U_0(x), V_0(x))\in E_m|1, 0)\\&=P\left\{\left(\frac{U_0(x)}{\sigma},\ \frac{V_0(x)-\theta}{\sigma}\right)\in E_m|\sigma, m\right\}\\&=P\{U_0(x)/m\leqslant\sigma\leqslant mU_0(x),\ -\sigma m\\&\leqslant\theta-V_0(x)\leqslant\sigma m|\sigma, \theta)=1.\end{aligned}$$

于是更有

$$\begin{aligned}P\{U_0(x)/m\leqslant\sigma\leqslant mU_0(x),\ -m^2U_0(x)\\\leqslant\theta-V_0(x)\leqslant m^2U_0(x)|\sigma, \theta\}=1.\end{aligned}$$

所以有理由猜想(σ, θ)的优良估计应满足关系

$$
\begin{aligned}
&U_0(x)/m \leqslant U(x) \leqslant mU_0(x); \\
&|V(x)-V_0(x)| \leqslant m^2 U_0(x).
\end{aligned} \tag{4.4}
$$

上式即$(S(x),\ r(x)) \in E_{m^2}$. 现在来证明其优良性. 对$(\sigma,\ \theta)$的任一估计$(U(x),\ V(x))$或相应的$(S(x),\ r(x))$, 定义

$$
(S^*(x),\ r^*(x)) = \begin{cases} (S(x),\ r(x)), & (S(x),\ r(x)) \in E_{m^2}; \\ (S(x),\ r(x)) \text{ 到 } E_{m^2} \text{ 的边界距离最} \\ \text{短的点}, & (S(x),\ r(x)) \overline{\in} E_{m^2}. \end{cases}
$$

如图所示, 在$(S,\ r) \overline{\in} E_{m^2}$时, (S, r)分别属于 I～VIII 八种情况, 显见, 当

$$(S,\ r) \in \text{I、II、III},\quad r^* = m^2 \leqslant r$$

时, 有

$$rU_0(x) + V_0(x) \geqslant r^* U_0(x) + V_0(x) \geqslant m + V_0(x) \geqslant 0,$$

当$(S,\ r) \in$ III、IV、V、$S^* = m^2 \leqslant S$, 有

$$SU_0(x) \geqslant S^* U_0(x) \geqslant m \geqslant 1.$$

当$(S,\ r) \in$ V、VI、VII, $r^* = -m^2 \geqslant r$, 有

$$rU_0(x) + V_0(x) \leqslant r^* U_0(x) + V_0(x) \leqslant -m + V_0(x) \leqslant 0.$$

当$(S,\ r) \in$ VII、VIII、I 时, $S^* = m^{-2} \geqslant S$, 有

$$SU_0(x) \leqslant S^* U_0(x) \leqslant m^{-1} \leqslant 1.$$

再根据$W(t_1,\ t_2)$满足 3° 的假设, 故在各种情况下,

$$
\begin{aligned}
&W[SU_0(y),\ rU_0(y) + V_0(y)] \\
&\quad \geqslant W[S^* U_0(y),\ r^* U_0(y) + V_0(y)].
\end{aligned}
$$

因此, $\rho(S,\ r;\ \sigma,\ \theta;\ P,\ W) \geqslant \rho(S^*,\ r^*;\ \sigma,\ \theta;\ P,\ W)$, 这就证得了$(S^*,\ r^*)$估计的优良性. 由$\underline{V}$的定义知, 对任意的$\varepsilon > 0$及$T > 1$, 存在$(S,\ r)$及相应的$(S^*,\ r^*)$使得先验分布$G$在$E_T$上取$(4T \log T)^{-1} d\theta\, d\sigma/\sigma$时,

$$
\begin{aligned}
\underline{V} + \varepsilon &\geqslant (4T \log T)^{-1} \int_{E_T} \rho(S,\ r;\ \sigma,\ \theta;\ P,\ W)\, d\theta\, d\sigma/\sigma \\
&\geqslant (4T \log T)^{-1} \int_{E_T} \rho(S^*,\ r^*;\ \sigma,\ \theta;\ P,\ W)\, d\theta\, d\sigma/\sigma
\end{aligned}
$$

$$=(4T\log T)^{-1}\int dP\int_{E_T} W(\cdot,\ \cdot)d\theta\, d\sigma/\sigma. \tag{4.5}$$

上式 $W(\cdot,\ \cdot)$ 中变量为 $(S^*(\sigma y+\theta 1^\tau)U_0(y),\ r^*(\sigma y+\theta 1^\tau)U_0(y)+V_0(y))$，我们又可得

$$\begin{aligned}
&\int_{E_T} W(\cdot,\ \cdot)d\theta\, d\sigma/\sigma \\
&\geqslant \int_{T^{-1}}^{T}\frac{d\sigma}{\sigma}\int_{-T-\sigma V_0(y)}^{T-\sigma V_0(y)} W(\cdot,\ \cdot)d\theta-2TMV_0(y) \\
&\geqslant \int_{U_0^{-1}(y)T^{-1}}^{U_0^{-1}(y)T}\frac{d\sigma}{\sigma}\int_{-T-\sigma V_0(y)}^{T-\sigma V_0(y)} W(\cdot,\ \cdot)d\theta-2TMV_0(y) \\
&\quad -2TM|\log U_0(y)| \\
&\geqslant \int_{E_y} W(\cdot,\ \cdot)d\theta\, d\sigma/\sigma-4TMm.
\end{aligned} \tag{4.6}$$

这里 $\quad E_y=\{(\sigma,\ \theta):(\sigma U_0(y),\ \sigma V_0(y)+\theta)\in E_T\}$;

$$M=\sup_{(t_1,t_2)\in E_{(m+1)^2}} W(t_1,\ t_2)$$

$$\geqslant \sup_{(\sigma,\theta)} W(S^*(\sigma y+\theta 1^\tau)U_0(y),\ r^*(\sigma y+\theta 1^\tau)U_0(y)+V_0(y))$$

其中用到了(4.3)的事实. 再作变换 $a=\sigma U_0(y)$, $b=aU_0(y)+\theta$, 于是

$$\int_{E_y} W(\cdot,\ \cdot)d\theta\, d\sigma/\sigma$$

$$=\int_{E_T} W(S^*_{a,b}(y),\ r^*_{a,b}(y))db\, da/a.$$

这里 $\quad S^*_{a,b}(y)=S^*\left(\frac{ay}{U_0(y)}+\left(b-\frac{aV_0(y)}{U_0(y)}\right)1^\tau\right)$;

$$r^*_{a,b}(y)=r^*\left(\frac{ay}{U_0(y)}+\left(b-\frac{aV_0(y)}{U_0(y)}\right)1^\tau\right).$$

对于固定的 a、b 都是不变统计量，它对应的 $\overline{U}^*(y)$、$\overline{V}^*(y)$ 是同变估计.

注意:

$$(4T\log T)^{-1}\int dP\int_{E_y} W(\cdot,\ \cdot)d\theta\, d\sigma/\sigma$$

$$\geqslant (4T\log T)^{-1}\int dP\int_{E_r} W(S^*_{a,b}(y)U_0(y),$$
$$r^*_{a,b}(y)U_0(y)+V_0(y))\,db\,da/a$$
$$\geqslant (4T\log T)^{-1}\int_{E_r} db\,da/a$$
$$\times\int W(S^*_{a,b}(y)U_0(y),\ r^*_{a,b}(y)U_0(y)+V_0(y))\,dP$$
$$\geqslant \inf_{(a,b)}\int W(S^*_{a,b}(y)U_0(y),\ r^*_{a,b}(y)U_0(y)+V_0(y))\,dP$$
$$\geqslant \inf_{(S,r\in\mathscr{I})}\int W(S(y)U_0(y),\ r(y)U_0(y)+V_0(y))\,dP=V. \quad (4.7)$$

由(4.5)、(4.6)、(4.7)知

$$\underline{V}+\varepsilon\geqslant V-TMm/T\log T\to V\ (\text{当}\ T\to\infty).$$

再由 ε 任意小，故 $\underline{V}\geqslant V$. 另一方面，又易得 $V\geqslant\overline{V}\geqslant\underline{V}$, 故 $\underline{V}=V$. 至此引理得证.

引理4.2 1) 设 $W\geqslant 0$ 在 Θ 上连续，$W(0,\ t_2)=\infty$, $W(t_1,\ t_2)\to\infty$ 对 $t_1^{-1}+t_1+|t_2|\to\infty$ 时一致成立.

μ 为定义在 Θ 上 Borel-σ 域上的测度，且对有界 Borel 集 B, $\mu(B)<\infty$, 则

$$\lim_{m\to\infty}\inf_{(a,b)\in\Theta}\int_{E_m} W(at_1,\ bt_1+t_2)\,d\mu$$
$$=\inf_{(a,b)\in\Theta}\int W(at_1,\ bt_1+t_2)\,d\mu. \quad (4.8)$$

2) 设 $W\geqslant 0$ 在 Θ 上连续，$W(t_1,\ t_2)\to\alpha<\infty$. 当 $t_1+|t_2|\to\infty$ 一致成立，则(4.8)式仍然成立.

证明 我们仅证明 1); 2)的证明是同样的，故省略. 不失一般性，m 取自然数，令

$$b_m(a,\ b)=\int_{E_m} W(at_1,\ bt_1+t_2)\,d\mu,$$

容易验证 $b_m(a,\ b)$ 在 Θ 上连续. 不失一般性，可设 $\mu(\Theta)>0$, 从而存在 m_0, 当 $m>m_0$ 时，有 $\mu(E_m)>0$. 由于假设 $W(t_1,\ t_2)\to\infty$, 当 $t_1^{-1}+t_1+|t_2|\to\infty$ 时一致成立. 因此 $h_m(a,\ b)$ 只能在有限

点$(a_m, b_m)\in\Theta$上达到极小. 记$C_m=h_m(a_m, b_m)$，显然它是m的单调非降函数. 极限$C_m\to C\leqslant\infty(m\to\infty)$存在. 记

$$h(a, b)=\int_\Theta W(at_1, bt_1+t_2)d\mu.$$

显然有

$$C=\lim_{m\to\infty}C_m\leqslant\inf_{(a,b)\in\Theta}h(a, b). \tag{4.9}$$

若$C=\infty$，对(4.8)式已对，若$C<\infty$，则首先可断言$\{a_m^{-1}+a_m+|b_m|\}$为有界，否则，若它为无界序列，不妨可设它趋于无穷，当$m\geqslant m_0$，$\mu(E_m)>0$，使用 Fatou 引理可得

$$\infty>C=\lim_{m\to\infty}C_m=\lim_{m\to\infty}h_m(a_m, b_m)$$

$$\geqslant\int_{E_{m_0}}\liminf_{m\to\infty}W(a_mt, b_mt+t_2)d\mu=\infty.$$

这就出现了矛盾. 故(a_m, b_m)有界，不妨设$(a_{m_k}, b_{m_n})\to(a_0, b_0)\in\Theta$，对任意确定的$T$，当$m_k\geqslant T$时，有

$$\int_{E_T}W(a_{m_k}t_1, b_{m_k}t_1+t_2)d\mu\leqslant\int_{E_{m_K}}W(a_{m_k}t_1, b_{m_k}t_1+t_2)d\mu$$

$$=C_{m_k}\leqslant C<\infty.$$

先令$m_k\to\infty$，再令$T\to\infty$，由上式就得到$h(a_0, b_0)\leqslant C$. 再对照(4.9)式，即可知(4.8)式成立，从而引理的1)成立.

引理 4.3 若μ满足引理 4.2 的要求，而且

$$d\mu=g(t_1, t_2)dt_1dt_2,$$

则当W在有界区域E_m上有界，引理 4.2 的结论仍成立.

证明 如同证明引理 3.3 那样，用多维 Ruzin 定理可以证得$h_m(a, b)$是a、b的连续函数，再用引理 4.2 的证法，即证得本引理.

引理 4.4 令$P(Z, B)$为$\mathscr{D}\times\mathscr{B}$上已知$Z$，$B\in\mathscr{B}$的条件概率，又设引理 4.2 及引理 4.3 的假设成立，则对任意$0<m<\infty$，有 Borel 可测函数$(a_m(Z), b_m(Z))$及$(a(Z), b(Z))$，使得

$$\int_{E_m}W(a_m(Z)t_1, b_m(Z)t_1+t_2)P(Z, d\omega)$$

$$=\inf_{(a,b)\in\Theta}\int_{E_m}W(at_1, bt_1+t_2)P(Z, d\omega), \tag{4.10}$$

$$\int W(a(Z)t_1,\ b(Z)t_1+t_2)P(Z,\ d\omega)$$

$$=\inf_{(a,b)\in\Theta}\int W(at_1,\ bt_1+t_2)P(Z,\ d\omega). \qquad (4.11)$$

此处 $d\omega$ 系指对 $(t_1,\ t_2)$ 的测度而言.

证明完全与引理 3.4 相同. 故从略.

定理 4.1 若损失函数满足 $1^\circ\sim3^\circ$ 及 4° 中之一(若 P 对 L 测度有密度, 可不要求 $W(t_1,\ t_2)$ 连续), 又 $P(A|0,\ 1)=0$, 则存在位置、刻度参数的最优同变估计, 而且还是 minimax 估计.

证明 在 R^n 定义函数

$Z(x)=(i_x,\ \mathrm{Sgn}(x_{i_x}-x_1),\ (x_{i_x}-x_j)/(x_{i_x}-x_1),\ j=2,\ \cdots,\ n)$, $i_x=\min\{j|x_j\neq x_1,\ j=2,\ \cdots,\ n\}$, $Z(x)$ 显然是在仿射群 $\mathscr{G}$ 下的最大不变统计量, $Z(x)$ 是 Borel 可测函数. 因此若 $(S,r)\in\mathscr{I}$, 则 S、r 可表为 $S(Z(x))$、$r(Z(x))$.

令 $P(Z,\ B)$ 是已知 Z、$(U_0(x),\ V_0(x))\in B$ 的条件概率. 固定 Z 后, 它是概率测度; 固定 B 后, 它是 $Z(x)$ 的 Borel 可测函数. 设 $(S,\ r)\in\mathscr{I}$, 则有

$$E_P\{W[S(X)U_0(X),\ r(X)U_0(X)+V_0(X)]|Z\}$$
$$=E_P\{W[S(Z)U_0(X),\ r(Z)U_0(X)+V_0(X)]|Z\}$$
$$=\int_\Theta W[S(Z)t_1,\ r(Z)t_1+t_2]P(Z,\ d\omega).$$

根据引理 4.4, 存在 Borel 可测函数 $(S_0,\ r_0)$, 使得

$$\int_\Theta W[S_0(Z)t_1,\ r_0(Z)t_1+t_2]P(Z,\ d\omega)$$
$$=\inf_{(a,b)\in\Theta}\int_\Theta W(at_1,\ bt_1+t_2)P(Z,\ d\omega).$$

因此对任意的 $(S,\ r)\in\mathscr{I}$, 我们有

$$\rho(S,\ r)=E_P\left\{\int W[S(Z)t_1,\ r(Z)t_1+t_2]P(Z,\ d\omega)\right\}$$
$$\geqslant E_P\left\{\int W[S_0(Z)t_1,\ r_0(Z)t_1+t_2]P(Z,\ d\omega)\right\}$$
$$=\rho(S_0,\ r_0)=V=V(P,\ W).$$

其次对正整数 m, 定义测度

$$P_m(C)=P(C\cap\{x:(U_0(x),\ V_0(x))\in E_m\}).$$

令 $h(x,\ m)$ 为 $\{x:(U_0(x),\ V_0(x))\in E_m\}$ 的示性函数, 对

$$(S,\ r)\in\mathscr{I},$$

$$E_P\{h(x,\ m)W[S(Z)U_0(x),\ r(Z)U_0(x)+V_0(x)]\,|\,Z\}$$

$$=\int_{E_m}W[S(Z)t_1+r(Z)t_1+t_2]P(Z,\ d\omega).$$

由引理 4.4, 存在 Borel 可测函数 $(S_m(Z),\ r_m(Z))$, 使得

$$\overline{W}_m(Z)=\int_{E_m}W[S_m(Z)t_1,\ r_m(Z)t_1+t_2]P(Z,\ d\omega)$$

$$=\inf_{(a,b)\in\Theta}\int_{E_m}W(at_1,\ bt_1+t_2)P(Z,\ d\omega).$$

与证明 $V=\rho(S_0,\ r_0)$ 同理可得

$$V_m=V(P_m,\ W)=\rho(S_m,\ r_m,\ P_m,\ W)=E_P(\overline{W}_m(Z)).$$

根据引理 4.2, $\overline{W}_m(Z)\to\overline{W}(Z)$, 当 $m\to\infty$ 时 $\overline{W}_m(Z)$ 是m的单调非降序列, 由单调收敛定理, $V(P_m,\ W)\to V(P,\ W)$ $(m\to\infty)$. 但我们已有 $V(P_m,\ W)=\underline{V}(P_m,\ W)\leqslant\underline{V}(P,\ W)$, 由引理 4.1 及 $\underline{V}\geqslant V=\underline{V}$, 所以

$$(U(x),\ V(x))=(U_0(x)S_0(x),\ U_0(x)r_0(x)+V_0(x))$$

是$(\sigma,\ \theta)$的最优同变 minimax 估计. 于是定理得证.

利用这个定理立即得到在 § 2 问题 3 中, 由 (2.12) 与 (2.13) 所确定的 θ、σ 的最优同变估计在所述平方损失下皆是 minimax 估计.

§ 5* 同变性与充分性、不变性与充分性

本节讨论同变性、不变性与充分性的关系. 众所周知, 依赖于充分统计量的随机化估计组成基本完全类, 进一步假设损失函数是凸的, 则依赖于充分统计量的非随机化估计组成基本完全类. 这自然产生一个问题: 依赖于充分统计量的同变估计类是否也是同变估计的基本完全类呢? 与此类似的问题是: 依赖于充分统计量的最大不变统计量, 在不变统计量中是否充分的? 这是

两个基本理论问题，本节介绍我国数理统计工作者[3]及 Stein 的有关结果.

本节假设样本空间为 $(\mathscr{X}, \mathscr{B}_{\mathscr{X}}, \mathscr{P})$，$\mathscr{P}=\{P_\theta, \theta\in\Theta\}$，$\mathscr{G}$ 为样本空间变换群，由 $\mathscr{G}$ 导出参数空间变换群 $\bar{\mathscr{G}}$，在行动空间导出变换群 $\mathscr{G}^*$，$\bar{\mathscr{G}}$、$\mathscr{G}^*$ 分别与 $\mathscr{G}$ 同态，在这些变换群下，统计判决问题不变.

现设 $t(x)$ 是 $(\mathscr{X}, \mathscr{B}_{\mathscr{X}}, \mathscr{P})$ 的一个充分统计量，其值域空间为 $(\mathscr{T}, \mathscr{B}_{\mathscr{T}})$，由 t 导出的分布族记为 $\mathscr{P}^T=\{P_\theta^T, \theta\in\Theta\}$.

为使依赖于 t 的同变估计类有意义，必须假定对每个 $g\in\mathscr{G}$，在 t 的值域 $\mathscr{T}$ 导出一个一一对应变换，即要求 $t(\mathscr{X})=\mathscr{T}$，且

$$t(x_1)=t(x_2)\Rightarrow t(gx_1)=t(gx_2) \quad 对\ \forall g\in\mathscr{G}\ 成立. \tag{5.1}$$

在此假定下，对任意的 $g\in\mathscr{G}$，在 $\mathscr{T}$ 相应导出一个变换 $\underline{g}$ 如下：任取 $t_0\in\mathscr{T}$，存在 $x_0\in\mathscr{X}$，$t(x_0)=t_0$，有

$$\underline{g}t_0=\underline{g}t(x_0)=t(gx_0).$$

引理 5.1 在上述假定下定义的 $\underline{g}$ 有

1°. $\underline{g}$ 是 $\mathscr{T}$ 到 $\mathscr{T}$ 的一一对应变换;

2°. $\underline{g^{-1}}=(\underline{g})^{-1}$，$\underline{g_1g_2}=\underline{g_1}\cdot\underline{g_2}$;

3°. $\underline{\mathscr{G}}=\{\underline{g}: g\in\mathscr{G}\}$ 是与 $\mathscr{G}$ 同态的群.

证明 1°. 任取 $t_0\in\mathscr{T}$，找到 x_0 使 $t(x_0)=t_0$，记 $x_0'=g^{-1}x_0$，$t_0'=t(x_0')$，则 $\underline{g}t_0'=t(gx_0')=t(x_0)=t_0$，因而知 $\underline{g}\mathscr{T}=\mathscr{T}$. 再令 t_0、$t_0'\in\mathscr{T}$，有 x_0、x_0' 使 $t(x_0)=t_0$，$t(x_0')=t_0'$. 若 $t(gx_0)=t(gx_0')$，由(5.1)知 $t(g^{-1}gx_0)=t(g^{-1}gx_0)$，即 $t_0=t(x_0)=t(x_0')=t_0'$，故 $\underline{g}$ 是 $\mathscr{T}$ 到 $\mathscr{T}$ 的一一对应变换.

2°. 任取 $t_0\in\mathscr{T}$，$(\underline{g^{-1}})t_0=t(g^{-1}x_0)=t_1$，存在 x_0、x_1 使 $t(x_0)=t_0$，$t(x_1)=t_1$. 即 $t(g^{-1}x_0)=t_1=t(x_1)$，由假设(5.1)知

$$t_0=t(x_0)=t(gg^{-1}x_0)=\underline{g}t(g^{-1}x_0)=\underline{g}(\underline{g^{-1}})t_0.$$

由 t_0 的任意性知 $\underline{g^{-1}}=(\underline{g})^{-1}$. 对任意的 $t_0=t(x_0)\in\mathscr{T}$，有 $\underline{g_1g_2}t_0=t(g_1g_2x_0)=\underline{g_1}t(g_2x_0)=\underline{g_1}\cdot\underline{g_2}t(x_0)=\underline{g_1}\cdot\underline{g_2}t_0$ 对 $\forall t_0\in\mathscr{T}$ 成立，故而 $\underline{g_1}\cdot\underline{g_2}=\underline{g_1g_2}$. 证得 2°. 而 3° 是 2° 的直接推论.

为了讨论以后的问题，有必要引进测度论中的 Blackwell 定理及其必要的概念.

定义 5.1 $A\subset R^1$ 称为 R^1 中的**解析集**，如果它是 R^2 中一个 Borel 集的投影集，即存在 R^2 Borel 集 B，使

$$A=\bigcup_y\{x: (x, y)\in B\}.$$

定义 5.2 可测空间 $(\mathscr{X}, \mathscr{B}_{\mathscr{X}})$ 称为 Blackwell 空间，如果 $\mathscr{B}_{\mathscr{X}}$ 可分(即由可数个集合生成的最小 σ 域)，且对 $(\mathscr{X}, \mathscr{B}_{\mathscr{X}})\to(\mathscr{T}, \mathscr{B}_{\mathscr{T}})$ 的每个可测函数 f 和每个 $A\in\mathscr{B}_{\mathscr{X}}$，$f(A)$ 是 R^1 中的解析集.

定义 5.3 $(\mathscr{X}, \mathscr{B}_{\mathscr{X}})$ 是可测空间，对 $\forall X\in\mathscr{X}$，定义

$$X = \bigcap_{x\in B,\, B\in \mathscr{B}_{\mathscr{X}}} B.$$

此集称为$\mathscr{B}_{\mathscr{X}}$的一个原子．若$T$是$(\mathscr{X}, \mathscr{B}_{\mathscr{X}})\to(\mathscr{T}, \mathscr{B}_{\mathscr{T}})$统计量，显然，$X_0=\{x:T(x)=T(x_0)\}$是$T^{-1}(\mathscr{B}_{\mathscr{T}})$的原子，$\forall x_0\in\mathscr{X}$．

在上述定义下，有如下两个重要结论：

引理5.2 设$(\mathscr{X}, \mathscr{B}_{\mathscr{X}})$为欧氏空间，$\mathscr{B}_{\mathscr{X}}$为 Borel 域，则$(\mathscr{X}, \mathscr{B}_{\mathscr{X}})$是 Blackwell 空间．

引理5.3 （Blackwell 定理）设$(\mathscr{X}, \mathscr{B}_{\mathscr{X}})$是 Blackwell 空间$\mathscr{B}_1$、$\mathscr{B}_2$是$\mathscr{B}_{\mathscr{X}}$的子$\sigma$域，且$\mathscr{B}_2$可分，则

$$\mathscr{B}_1\subseteq\mathscr{B}_2\Leftrightarrow\mathscr{B}_1\text{的原子是}\mathscr{B}_2\text{原子的并}.$$

这两个引理的证明见[1]，此处从略．

引理5.4 假定

1) $(\mathscr{X}, \mathscr{B}_{\mathscr{X}})$为欧氏空间，$\mathscr{B}_{\mathscr{X}}$为 Borel 集组成的．

2) T为$(\mathscr{X}, \mathscr{B}_{\mathscr{X}}, \mathscr{P})$的充分统计量，$\mathscr{B}_{\mathscr{T}}$可分，且包含$\mathscr{T}$的任一单点作为原子．

则对任意的$g\in\mathscr{G}$，有

$$gT^{-1}(\mathscr{B}_{\mathscr{T}})=T^{-1}(\mathscr{B}_{\mathscr{T}}), \tag{5.2}$$

$$\underline{g}\mathscr{B}_{\mathscr{T}}=\mathscr{B}_{\mathscr{T}}, \tag{5.3}$$

$$P_{g\theta}^{T}(\underline{g}B)=P_{\theta}^{T}(B)\quad \text{对}\ \forall\theta\in\Theta,\ B\in\mathscr{B}_{\mathscr{T}}. \tag{5.4}$$

证明 首先验证下面的等式：对$\forall g\in\mathscr{G}$, $B\in\mathscr{B}_{\mathscr{T}}$,

$$T^{-1}(\underline{g}B)=gT^{-1}(B),\ T(gT^{-1}B)=\underline{g}(B). \tag{5.5}$$

事实上，$x\in T^{-1}(\underline{g}B)\Leftrightarrow T(x)\in\underline{g}B\Leftrightarrow\underline{g}^{-1}T(x)\in B\Leftrightarrow T(g^{-1}x)\in B\Leftrightarrow g^{-1}x\in T^{-1}(B)\Leftrightarrow x\in gT^{-1}(B)$．

又$t_0\in T(gT^{-1}B)\Leftrightarrow$存在$x_0\in T^{-1}(B)$使$t_0=T(gx_0)\Leftrightarrow$存在$x_0$, $T(x_0)\in B$, $\underline{g}T(x_0)=t_0\Leftrightarrow\underline{g}^{-1}t_0\in B\Leftrightarrow t_0\in\underline{g}(B)$．从而证得(5.5)式．

现在考察由$(\mathscr{X}, B_{\mathscr{X}})$到$(\mathscr{T}, \mathscr{B}_{\mathscr{T}})$的两个变换$T(x)$及$T(g^{-1}x)$，它们在$(\mathscr{X}, B_{\mathscr{X}})$中导出的$\sigma$域分别为

$$\mathscr{B}_1=T^{-1}(\mathscr{B}_{\mathscr{T}}),\ \mathscr{B}_2=gT^{-1}(\mathscr{B}_{\mathscr{T}}).$$

由于g和T在各自意义下的可测性，故$\mathscr{B}_1$与$\mathscr{B}_2$都是$\mathscr{B}_{\mathscr{X}}$的子$\sigma$域．又$\mathscr{B}_{\mathscr{T}}$可分，故$\mathscr{B}_1$、$\mathscr{B}_2$皆可分．又因$\mathscr{B}_{\mathscr{T}}$包含$\mathscr{T}$的一切单点集，故$\mathscr{B}_2$的任一原子有$\{x:T(g^{-1}x)=t_0\}$ $(t_0\in\mathscr{T})$的形式，$\mathscr{B}_1$的任一原子有$\{x:T^{-1}(x)=t_0\}$ $(t_0\in\mathscr{T})$的形式．记$t_1=\underline{g}t_0$，易见$\{x:T(g^{-1}x)=t_0\}=\{x:T(x)=t_1\}$，于是$\mathscr{B}_2$的每个原子必为$\mathscr{B}_1$的某个原子，根据对$(\mathscr{X}, \mathscr{B}_{\mathscr{X}})$的假定和 Blackwell 定理知$\mathscr{B}_1\subseteq\mathscr{B}_2$，由$g$的任意性及$\mathscr{G}$为群知$\mathscr{B}_2\subseteq\mathscr{B}_1$也成立，即$\mathscr{B}_1=\mathscr{B}_2$，(5.2)成立．

现取$B\in\mathscr{B}_{\mathscr{T}}$，有$gT^{-1}(B)\in gT^{-1}(\mathscr{B}_{\mathscr{T}})=T^{-1}(\mathscr{B}_{\mathscr{T}})$，故存在$B_1\in\mathscr{B}_{\mathscr{T}}$，使

$T^{-1}(B_1)=g_T^{-1}(B)$，再由(5.5)得 $B_1=T(T^{-1}(B_1))=T(gT^{-1}(B))=\underline{g}(B)$，因而 $\underline{g}(B)\in\mathscr{B}_{\mathscr{T}}$. 由于 $\underline{g}$ 为群. 这就证明了(5.2).

最后由(5.5)知,

$$P_{\underline{g}\theta}^T(\underline{g}B)=P_{\bar{g}\theta}(gT^{-1}(B))=P_\theta(T^{-1}B)=P_\theta^T(B),$$

故(5.3)式得证,至此引理证完.

引理5.5 设引理 5.1 及引理 5.4 条件成立, T 为充分统计量, 记 $P(\cdot|t)$ 为给定 $T=t$ 时 x 的正则条件概率函数, 则它是几乎同变的, 即对 $\forall g\in\mathscr{G}$, 存在 $N_g\in\mathscr{B}_{\mathscr{T}}$, $\mathscr{P}^T(N_g)=0$, 且

$$P(gC|\underline{g}t)=P(C|t)\quad 对\ \forall C\in B_{\mathscr{X}},\ t\bar{\in} N_g. \tag{5.6}$$

证明 考察定义于 $\mathscr{B}_{\mathscr{X}}\times\mathscr{T}$ 上的函数 $Q_g(C|t)=P(gC|\underline{g}t)$, 显然, 给定 t 时, $Q_g(\cdot|t)$ 为 $\mathscr{B}_{\mathscr{X}}$ 上的概率测度,由引理5.4知给定 C 时, $Q_g(C|t)$ 为 $\mathscr{B}_{\mathscr{T}}$ 可测. 根据(5.4)与(5.5), 可证明对 $\forall B\in\mathscr{B}_{\mathscr{X}}$, $C\in\mathscr{B}_{\mathscr{X}}$, 有

$$\begin{aligned}\int_B Q_g(C|t)P_\theta^T(dt)&=\int_B P(gC|\underline{g}t)P_{\bar{g}\theta}^T(d\underline{g}t)\\&=P_{\bar{g}\theta}(gC\cap T^{-1}(\underline{g}B))=P_\theta(g^{-1}(gC\cap T^{-1}(\underline{g}B))\\&=P_\theta(C\cap g^{-1}T^{-1}(\underline{g}B))=P_\theta(C\cap T^{-1}(B)).\end{aligned}$$

即 $Q_g(\cdot|\cdot)$ 也是给定 t 时 x 的正则条件概率函数. 故对 $\forall C\in B_{\mathscr{X}}$, 存在 $N_{g,C}$, 使 $\mathscr{P}^T(N_{g,C})=0$, 且

$$当\ t\bar{\in} N_{g,C}\ 时,\quad P(gC|\underline{g}t)=P(C|t).$$

将一切形如 $\{x:x<r\}$ 的集(其中 r 为 $\mathscr{X}$ 中的有理点)排列为 $C_1, C_2, \cdots$, 令 $N_g=\bigcup\limits_{i=1}^{\infty}N_{g,C_n}$, 则 $\mathscr{P}^T(N_g)=0$, 且

$$当\ t\bar{\in} N_g\ 时,\quad P(gC_n|\underline{g}t)=P(C_n|t)(n=1,\ 2,\ \cdots).$$

由于 $P(gC|\underline{g}t)$ 和 $P(C|t)$ 都是 $\mathscr{B}_{\mathscr{X}}$ 上的概率测度, 由测度扩张唯一性定理推得(5.6)成立. 从而引理得证.

上述几个引理可使我们断言: 依赖于充分统计量的几乎同变随机化估计组成几乎同变估计中的基本完全类. 但为了进一步取消"几乎"二字, 同时也为建立几乎同变估计与同变估计的联系,需加强条件作进一步讨论.

引理5.6 假设

i) $(\mathscr{X},\ \mathscr{B}_{\mathscr{X}})$ 为可分空间,且单点集可测.

ii) $\mathscr{G}$、$\underline{\mathscr{G}}$ 分别为 $(\mathscr{X},\ \mathscr{B}_{\mathscr{X}})$、$(\mathscr{T},\ \mathscr{B}_{\mathscr{T}})$ 上的可测变换群,且 $\underline{\mathscr{G}}$ 与 $\mathscr{G}$ 同态. 在 $\mathscr{G}$ 上也引进 σ-域结构 $\mathscr{B}_g$ 及其上的 σ 有限测度 ν 满足条件: $\nu\neq0$, 对任意的 $G\in\mathscr{B}_g$, 有 $G_g\in\mathscr{B}_g$ 及 $\nu(G)=\nu(G_g)$ 关于 $\forall g\in\mathscr{G}$ 成立.

iii) $\delta(gC|\underline{g}t)$ 对 $\forall C\in\mathscr{B}_{\mathscr{X}}$ 是关于 $\mathscr{B}_g\times\mathscr{B}_{\mathscr{T}}$ 可测,且满足

$$\delta(gC|\underline{g}t)=\delta(C|t)\quad 对\ \forall t\bar{\in} N_g\ 及\ C\in\mathscr{B}_{\mathscr{X}}.$$

这里 $\mathscr{P}^T(N_g)=0$, N_g 与 C 无关,即 δ 是几乎同变估计.

则存在 $\delta^*(\cdot|t)$是关于 $\mathscr{G}$ 同变的,即对于 $\forall t\in\mathscr{T}$、$C\in\mathscr{B}_{\mathscr{X}}$、$g\in\mathscr{G}$, $\delta^*(gC|\underline{g}t)=\delta^*(C|t)$. 而且存在 $\mathscr{P}$ 零测集 N_0, 当 $t\bar{\in} N_0$ 时有 $\delta^*(C|t)=\delta(C|t)$ 对 $\forall C\in\mathscr{B}_{\mathscr{X}}$ 成立.

证明 考虑函数

$$Q_n(g,\ t)=\delta(gC_n|\underline{g}t)-\delta(C_n|t). \tag{5.7}$$

这里 $\{C_n\}$ 的意义与引理5.5中的相同. 由 iii)知: $Q_n(g,\ t)$ 关于 $\mathscr{B}_g\times\mathscr{B}_{\mathscr{T}}$ 可测,于是

$$E_n\triangleq\{(g,\ t),\ Q_n(g,\ t)\neq 0\}\in\mathscr{B}_g\times\mathscr{B}_{\mathscr{T}}.$$

对固定的 g, 由于 $\mathscr{P}^T\{t:(g,\ t)\in E_n\}=0$, 记

$$N_n=\{t:\nu(g:(g,\ t)\in E_n)\neq 0\},\ N=\bigcup_{n=1}^{\infty}N_n.$$

运用 Fubini 定理, 即得 $\mathscr{P}^T(N_n)=\mathscr{P}^T(N)=0$, 又由于 $\nu\not\equiv 0$, ν 为 σ 有限测度,不失一般性, 可设 $\nu(\mathscr{G})=1$. 例如令 $g=\bigcup_{i=1}^{\infty}g_i$, $\{g_i\}$ 彼此不交, $g_i\in\mathscr{B}_g$, $\nu(g_i)>0$, 对 $G\in\mathscr{B}_g$, 令 $\bar{\nu}(G)=\sum_{i=1}^{\infty}\nu(g_i\cap G)/2^i\nu(g_i))$, 而后定义

$$f(g,\ t)=\sum_{n=1}^{\infty}\left|\int_g\delta(gC_n|\underline{g}t)d\nu(g)-\delta(gC_n|\underline{g}t)\right|,$$

$$R=\left\{t:\int_g f(g,\ t)d\nu(g)=0\right\}.$$

则易知 $R\in\mathscr{B}_{\mathscr{T}}$. 我们先指出 $\mathscr{P}^T(R)=1$. 对任一 $t_0\bar{\in} N$, 由于 $\delta(gC_n|\underline{g}t_0)=\delta(C_n|t_0)$, a. s. $\nu(n=1,\ 2,\ \cdots)$. 由此 a. s. ν 成立着下列等式:

$$\int_g\delta(g'C_n|\underline{g}'t_0)d\nu(g')=\delta(gC_n|\underline{g}t_0)\quad(n=1,\ 2,\ \cdots). \tag{5.8}$$

从而显见 $f(g,\ t_0)=0$, a. s. ν. 即 $t_0\in R$, 说明 $R\supset\mathscr{T}-N$, 自然 $\mathscr{P}^T(R)=1$. 再由 $t_0\in R$ 可使(5.8)式 a. s. ν 成立, 而 $\mathscr{B}_{\mathscr{X}}$ 由 C_1, C_2, $\cdots$,所生成,利用 $(\mathscr{X},\ B_{\mathscr{X}})$可分及测度扩张定理知, 存在 ν 零测集 M, 只要 $g\bar{\in} M$, 对于任何 $C\in\mathscr{B}_{\mathscr{X}}$ 皆有

$$\int_g\delta(g'C|g't_0)d\nu(g')=\delta(gC|\underline{g}t_0). \tag{5.9}$$

现在再指出 R 为群 g 的不变集, 即对于任意的 $g_0\in\mathscr{G}$, $g_0R=R$, 当取 $t_0\in R$ 时, 证明 $t_1=g_0t_0\in R$. 对于任意的 $C\in B_{\mathscr{X}}$, 令 $C=g_0D$. 利用假设 ii) $d\nu(g)=d\nu(gg_0)$及(5.9)式,有

$$\int_g\delta(g'C|\underline{g}'t_1)d\nu(g')=\int_g\delta(g'g_0D|\underline{g'g_0}t_0)d\nu(g')$$

$$=\int_g\delta(g''D|\underline{g}''t_0)d\nu(g'')=\delta(gD|\underline{g}t_0)=\delta(gg_0^{-1}C|\underline{gg_0^{-1}}t_1) \tag{5.10}$$

对 $g\bar{\in}M$ 成立，然而，由 $g\bar{\in}M\Rightarrow gg_0^{-1}\bar{\in}Mg_0^{-1}$，同时由 $\nu(M)=0\Rightarrow\nu(Mg_0^{-1})=0$，所以(5.10)式左端就等于 $\delta(gC|\underline{g}t_1)$(当 $g\bar{\in}Mg_0^{-1}$)，再令 C 取 C_1, C_2, $\cdots$，即得 $f(g, t_1)=0$, a. s. ν. 所以 $t_1\in R$. 而 g_0 是群 $\mathscr{G}$ 的任一元素，所以 R 为不变集.

现在开始引进 δ^*，对于 $t\in R$，令

$$\delta^*(C|t)=\int_{\mathscr{G}}\delta(gC|\underline{g}t)d\nu(g)\quad(t\in R).$$

因为对于 $\forall g_0\in\mathscr{G}$, $C\in\mathscr{B}_{\mathscr{X}}$, $t_0\in R$, 由于 $d\nu(g)=d\nu(gg')$，故有

$$\begin{aligned}\delta^*(g_0C|\underline{g}_0t_0)&=\int\delta(g'g_0C|\underline{g}'\underline{g}_0t_0)d\nu(g')\\&=\int\delta(g''C|\underline{g}''t)d\nu(g'')=\delta^*(C|t_0).\end{aligned}$$

从而知: 对于 $t\in R$, δ^* 是同变的.

现在对于 $t\in\mathscr{T}-R$，构造 δ^*，先建立一个由 $\mathscr{T}-R$ 到 R 的映射(不必一一对应)如下: 任取 $t_0\in\mathscr{T}-R$, $t_0'\in R$. 令 $\underline{g}t_0$ 对应 $\underline{g}t_0'$, $\underline{g}\in\underline{\mathscr{G}}$. 如果 $\mathscr{T}-R-\{\underline{g}t_0, g\in\mathscr{G}\}$ 是非空的，则再取其中一个点 t_1，令 $\underline{g}t_1$ 对应 $\underline{g}t_0'$, $\underline{g}\in\underline{\mathscr{G}}$，依此作下去，必要时应用选择公理，可以构造出由 $\mathscr{T}-R$ 到 R 的一个映射，记之为 h. 若 $a\in\mathscr{T}-R$，显然，有 $h(\underline{g}a)=\underline{g}h(a)$. 再定义

$$\delta^*(C|t)=\delta^*(C|h(t))\quad(\text{当 } t\bar{\in}R).$$

注意: $h(t)\in R$，从而当 $t\bar{\in}R$ 时，有

$$\delta^*(gC|\underline{g}t)=\delta^*(gC|h(\underline{g}t))=\delta^*(gC|\underline{g}h(t))=\delta^*(C|h(t))=\delta^*(C|t).$$

这就得出 δ^* 对于 $t\bar{\in}R$ 时也是同变的. 由引理对 $\delta(\cdot|\cdot)$的假定，也易知，对于固定的 t, $\delta^*(C|t)$ 是 $\mathscr{B}_{\mathscr{X}}$ 上的概率测度. 固定 $C\in\mathscr{B}_{\mathscr{X}}$, $\delta^*(C|\cdot)$ 为 $\mathscr{B}_{\mathscr{T}}$ 可测函数.

最后，在(5.7)式两端对 ν 积分得: 当 $t\bar{\in}N$, $t\in R$时，有

$$\begin{aligned}0=\int Q_n(g', t)d\nu(g')&=\int\delta(g'C_n|\underline{g}'t)d\nu(g)-\delta(C_n|t)\\&=\delta^*(C_n|t)-\delta(C_n|t)\quad(n=1, 2, \cdots).\end{aligned}$$

而 $\mathscr{P}^T(N)=0$, $\mathscr{P}^T(R)=1$. 记 $N_0=N\cup(\mathscr{T}-R)$. 再由测度扩张定理知:

$$\delta^*(C|t)=\delta(C|t)\quad(\text{当 } t\bar{\in}N_0 \text{ 时}).$$

至此引理证完.

定理5.1 对于不变统计判决问题，假设引理5.5的条件成立. 行动空间为欧氏空间，$\mathscr{B}_D$ 为其中的 Borel 集所组成. 则依赖于充分统计量的几乎同变随机化估计是几乎同变随机化估计中的基本完全类. 若进一步假设引理5.6ii)成立，则依赖于充分统计量 T 的同变随机化估计是同变随机化估计类的基本完全类.

证明 设 δ 为几乎同变估计，以 $P(\cdot|t)$ 表示给定 $T=t$ 时 x 的任一正则条件概率函数. 定义依赖于 t 的判决函数

$$\bar{\delta}(D|t)=\int_{\mathscr{X}}\delta(D|x)P(dx|t),\ \forall D\in\mathscr{B}_{D},\ t\in\mathscr{T}.$$

众所周知，$\bar{\delta}$ 的风险函数与 δ 的风险函数相同，故证明了 $\bar{\delta}$ 为几乎同变估计后即得定理的前一结论. 任给 $g\in\mathscr{G}$，根据引理 5.5 的 (5.6) 式，存在 $N_g\in\mathscr{B}_{\mathscr{T}}$，$\mathscr{P}^T(N_g)=0$，使 $P(C|t)=P(g^{-1}C|\underline{g}^{-1}t)$（当 $t\bar{\in}N_g$，$\forall C\in\mathscr{B}_{\mathscr{X}}$ 时）. 再利用定理对 $(\mathscr{D},\ \mathscr{B}_{\mathscr{D}})$ 的假定，由 δ 是几乎同变估计知：存在集 $M_g\in\mathscr{B}_{\mathscr{X}}$，$\mathscr{P}(M_g)=0$，使

$$\delta(g^*A|gx)=\delta(A|x)\quad (\text{当 } x\bar{\in}M_g,\ A\in\mathscr{B}_{\mathscr{D}}).$$

记 $N_1=\{t:P(M_g|t)\neq0\}$，则 $N_1\in\mathscr{B}_{\mathscr{T}}$. 由 Fubini 定理可知 $\mathscr{P}^T(N_1)=0$，再令 $N_{1g}=N_1\cup N_g$，得 $\mathscr{P}^T(N_{1g})=0$，当 $t\bar{\in}N_{1g}$ 时，

$$\begin{aligned}\bar{\delta}(A|t)&=\int_{\mathscr{X}}\delta(A|g^{-1}x)P(dg^{-1}x|\underline{g}^{-1}\underline{g}t)\\&=\int_{\mathscr{X}}\delta(g^*A|x)P(dx|\underline{g}t)=\bar{\delta}(g^*A|\underline{g}t).\end{aligned}$$

因此，$\bar{\delta}$ 为几乎同变估计. 定理前半部分得证. 现在证明定理的后半部分：由假设 $(\mathscr{X},\ \mathscr{B}_{\mathscr{X}})$ 为欧氏空间可分，引理 5.6 的假设 i) 成立，本定理又设引理 5.6 的假设 ii) 成立. 再由引理 5.5 知，正则概率函数 $P(\cdot|\cdot)$ 是几乎同变的，符合引理 5.6 的假设 iii)，故有引理 5.6 结果，存在正则条件概率 $P(\cdot|t)$ 为同变，对任意同变估计 $\delta(\cdot|x)$，构造

$$\bar{\delta}(A|t)=\int_{\mathscr{X}}\delta(A|x)P(dx|t)\quad (A\in\mathscr{B}_{\mathscr{D}}).$$

易知 $\bar{\delta}(\cdot|t)$ 的风险函数与 $\delta(\cdot|x)$ 的风险函数相同. 且对 $\forall g\in\mathscr{G}$，有

$$\begin{aligned}\bar{\delta}(g^*A|\underline{g}t)&=\int_{\mathscr{X}}\delta(g^*A|gx)P(dgx|\underline{g}t)\\&=\int_{\mathscr{X}}\delta(A|x)P(dx|t)=\bar{\delta}(A|t).\end{aligned}$$

因此 $\bar{\delta}(\cdot|t)$ 也是同变估计. 定理得证.

注 如果每个依赖于充分统计量的几乎同变估计对任何 $A\in\mathscr{B}_{\mathscr{D}}$，关于 $\mathscr{B}_g\times\mathscr{B}_{\mathscr{T}}$ 可测，则在定理 5.1 的条件下，利用引理 5.6 即知：存在同变估计 $\delta^*(A|t)=\delta(A|t)$，a. s. $\mathscr{P}^T$. 对 A 一致成立. 这说明同变估计 $\delta^*(A|t)$ 类是几乎同变估计类的基本完全类.

在点估计中，最注目的还是非随机化估计，这时须更强的条件，才能使类似的定理成立.

定理 5.2 若定理 5.1 的条件成立，并假设损失函数是凸的（且 $L(\theta,\ d)\to\infty$，当 $\|d\|\to\infty$）. $\mathscr{D}$ 为有限维欧氏空间或其凸子集，$\mathscr{B}_{\mathscr{D}}$ 为 $\mathscr{D}$ 上的

Borel 集类. $\mathscr{G}$ 是线性变换群或有限群, 则依赖充分统计量 T 的非随机化几乎同变估计类在具有有限风险的非随机化几乎同变估计类中构成基本完全类.

证明 由引理 5.5 知, 存在几乎同变正则条件概率分布 $P(\cdot|t)$, 即对 $\forall g \in \mathscr{G}$, 存在 $\mathscr{P}^T$ 零测集 M_g, 使 $P(gx|\underline{g}t)=P(x|t)$, $t \overline{\in} M_g$. 若 $\delta(x)$ 是几乎同变估计, 且 $R(\theta, \delta)<\infty$, $\forall \theta \in \Theta$, 则由于 $L(\theta, d)$ 的凸性, 存在 $\alpha_\theta>0$ 及 β_θ, 使得 $L(\theta, d) \geqslant \alpha_\theta\|d\|+\beta_\theta$. 于是存在 $\mathscr{P}^T$ 零测集 N^T, 当 $t \overline{\in} N^T$ 时, 有

$$\infty>\int L(\theta, \delta(x)) P(dx|t) \geqslant \alpha(\theta) \int\|\delta(x)\| P(dx|t)+\beta(\theta),$$

即 $\int\|\delta(x)\| P(dx|t)<\infty$. 我们就定义

$$\delta^*(t)=\begin{cases}\int \delta(x) P(dx|t), & \text{当 } t \overline{\in} N^T; \\ a, & \text{当 } t \in N^T.\end{cases}$$

式中 $a \in \mathscr{D}$, 易见对于 $\forall g \in \mathscr{G}$, $t \overline{\in} N^T \cup M_g$ 有

$$\begin{aligned} g^* \delta^*(t) & =g^* \int \delta(x) P(dx|t)=\int g^* \delta(x) P(dx|t) \\ & =\int \delta(gx) P(dx|t)=\int \delta(x') P(dg^{-1}x'|t) \\ & =\int \delta(x') P(dx'|\underline{g}t)=\delta^*(\underline{g}t).\end{aligned}$$

故 $\delta^*(t)$ 是几乎同变的. 因损失函数为凸的. 由 Jensen 不等式就得到 $R(\theta, \delta^*) \leqslant R(\theta, \delta)$, $\forall \theta \in \Theta$.

现在证明同变估计的情形: 由于 $\mathscr{G}$ 的构造知引理 5.6 的假设 ii) 成立. 故存在同变正则条件概率 $P(x|t)$, 若 $\delta(x)$ 是同变非随机化估计, 存在 N^T, $\mathscr{P}^T(N^T)=0$,

$$\int\|\delta(x)\| P(dx|t)<\infty \quad (\text{当 } t \overline{\in} N^T).$$

于是可以构造

$$\delta^*(t)=\begin{cases}\int \delta(x) P(dx|t), & \text{当 } t \overline{\in} N^T; \\ a, & \text{当 } t \in N^T.\end{cases}$$

$a \in \mathscr{D}$, 对于 $g \in \mathscr{G}$, $t \overline{\in} N^T$, 有

$$g^* \delta^*(t)=\int g^* \delta(x) P(dx|t)=\int \delta(x') P(dx'|\underline{g}t).$$

这说明 $\underline{g}t \overline{\in} N^T$, 所以 $\mathscr{T}-N^T$ 是不变的.

当 $t \in N^T$, 利用引理 5.6 证明 $P(x|t)$ 同变的同样办法建立 N^T 与 $\mathscr{T}-N^T$ 的映射 $h(t)$, 使 $\underline{g}h(t)=h(\underline{g}t)$, 定义

$$\delta^*(t)=\delta^*(h(t)),\quad \text{当 } t\in N^T,\ h(t)\in\mathscr{T}-N^T.$$

从而

$$\delta^*(\underline{g}t)=\delta^*(h(\underline{g}t))=\delta^*(\underline{g}h(t))$$
$$=g^*\delta^*(h(t))=g^*\delta^*(t).$$

如此给出的 $\delta^*(t)$ 为同变的，由损失函数的凸性及 Jensen 不等式得出 $R(\theta,\delta^*)\leqslant R(\theta,\delta)$, $\theta\in\Theta$, 由此定理即得证.

下面讨论第二个问题：最大不变统计量 $I(x)$ 与充分统计量 $S(x)$ 的关系. 为此，要把充分 σ 域及充分统计量作点引伸：给定概率空间 $(\mathscr{X},\mathscr{B}_{\mathscr{X}},\mathscr{P})$, 且 $\mathscr{B}_1\subseteq\mathscr{B}_2\subseteq\mathscr{B}_{\mathscr{X}}$. 如果对 $\forall A\in\mathscr{B}_2$, 存在一个现实 $E(I_A(x)|\mathscr{B}_1)$ 与 $P\in\mathscr{P}$ 无关. 则称 $\mathscr{B}_1$ 关于$(\mathscr{X},\mathscr{B}_2,\mathscr{P})$**是充分的**.

令 $\mathscr{B}_I$ 是最大不变 σ 域, $\mathscr{B}_I\subseteq\mathscr{B}_{\mathscr{X}}$. 一个不变统计量 $S_I(x)$ 称为**不变充分统计量**，若 $S_I(x)$ 产生的 σ 域 $\mathscr{B}_{SI}\subseteq\mathscr{B}_I$ 是关于$(\mathscr{X},\mathscr{B}_I,\mathscr{P})$充分的.

若 $S(x)$ 是 $\mathscr{P}$ 的充分统计量，它产生的 σ 域为 $\mathscr{B}_S\subseteq\mathscr{B}_{\mathscr{X}}$, 令 $\mathscr{B}_{SI}=\mathscr{B}_S\cap\mathscr{B}_I$. 那末，在什么条件下，$\mathscr{B}_{SI}$ 是关于$(\mathscr{X},\mathscr{B}_I,\mathscr{P})$充分的？下面介绍 Stein[16]、[17]的结果.

定理 5.3 设 $V=U(S)$ 为关于变换群 $\mathscr{G}$ 的基于 S 的最大不变统计量，S 为 $\mathscr{P}$ 的充分统计量，当下面两个条件之一成立时，则 V 也是不变充分统计量.

条件 A: i) 对于 $\forall g\in\mathscr{G}$, 有 $g\mathscr{B}_S=\mathscr{B}_S$.

ii) f_S 为 $\mathscr{B}_S$ 可测函数，关于 $\mathscr{G}$ 几乎不变，则存在 $\mathscr{B}_{SI}$ 上的可测函数 f_{SI}, 与 f_S 等价.

条件 B: 存在一个不变条件概率测度，即

$$P(B|S)=P(gB|\underline{g}S)\quad(\forall B\in\mathscr{B}_{\mathscr{X}},\ g\in\mathscr{G}).$$

下面通过一组引理来证明这个定理.

引理 5.7 若条件 A 成立，则 $\mathscr{B}_{SI}$ 对 $\mathscr{P}^I$ 充分.

证明 若 f_I 为不变统计量，P 可积，这里 $X\sim P$, 记 $gX\sim\overline{P}$, 对于 $A\in\mathscr{B}_S$, 在条件 A i)之下，有

$$\int_A E(f_I|S(gx))dP=\int_{gA}E(f_I|S(x))d\overline{P}(x)$$
$$=\int_{gA}f_I(x')d\overline{P}(x')=\int_A f_I(gx)dP$$
$$=\int_A f_I(x)dP=\int_A E(f_I|S(x))dP.$$

从而有

$$E(f_I|S(gx))=E(f_I|S(x))\quad \text{a. s. } \mathscr{P}^S,$$

即 $E(f_I|S(x))=E(f_I|\mathscr{B}_S)$ 对 $\mathscr{G}$ 几乎不变，在条件 A ii)之下，存在 $\mathscr{B}_{SI}$ 可测函数 $f_{SI}(x)$ 与 $E(f_I|\mathscr{B}_S)$ 等价.

$$E_P\{f_I(x)|\mathscr{B}_{SI}\}=E_P\{E_P(f_I(x)|\mathscr{B}_S)|\mathscr{B}_{SI}\}$$

故 $$=E_P(f_{SI}(x)|\mathscr{B}_{SI})=f_{SI}(x), \quad \text{a. s. } \mathscr{P}^{\mathrm{I}}.$$

上述等式右边与 P 无关，即 $\mathscr{B}_{SI}$ 对 $\mathscr{P}^{\mathrm{I}}$ 是充分的.

引理 5.8 下面三个事实是等价的:

i) 若 f_I 不变，则存在 $f_{SI}(x)$ 为 $\mathscr{B}_{SI}$ 可测函数，使

$$f_{SI}(x)=E(f_I|\mathscr{B}_S). \quad \text{a. s. } \mathscr{P}^S.$$

ii) 若 f_I 不变，则

$$E(f_I|\mathscr{B}_S)=E(f_I|\mathscr{B}_{SI}). \quad \text{a. s. } \mathscr{P}^S.$$

iii) 若 f_I 不变，$f_S(x)$ 为 $\mathscr{B}_S$ 可测函数，则

$$E\{f_I(x)f_S(x)|\mathscr{B}_{SI}\}=E(f_I(x)|\mathscr{B}_{SI})E(f_S(x)|\mathscr{B}_{SI}). \quad \text{a. s. } \mathscr{P}^S.$$

证明 ii)⇒i)是显然的；下面证明 i)⇒iii):

$$\begin{aligned}E\{f_I(x)f_S(x)|\mathscr{B}_{SI}\}&=E\{E(f_I(x)f_S(x)|\mathscr{B}_S)|\mathscr{B}_{SI}\}\\&=E\{f_S(x)E(f_I(x)|\mathscr{B}_S)|\mathscr{B}_{SI}\}\\&=f_{SI}(x)E(f_S(x)|\mathscr{B}_{SI})\\&=E(f_I|\mathscr{B}_{SI})E(f_S|\mathscr{B}_{SI}). \quad \text{a. s. } \mathscr{P}^S.\end{aligned}$$

再证 iii)⇒ii): 对任意的 $A\in\mathscr{B}_S$, 有

$$\begin{aligned}&\int_A E(f_I(x)|\mathscr{B}_S)dP=\int_A f_I(x)dP=\int I_A f_I(x)dP\\&=E[E(I_A(x)f_I(x)|\mathscr{B}_{SI})]=E[E(I_A|\mathscr{B}_{SI})E(f_I(x)|\mathscr{B}_{SI})]\\&=E[E\{I_A(x)E(f_I|\mathscr{B}_{SI})|\mathscr{B}_{SI}\}]=E[I_A E(f_I|\mathscr{B}_{SI})]\\&=\int_A E(f_I(x)|\mathscr{B}_{SI})dP\\&\Rightarrow E(f_I(x)|\mathscr{B}_S)=E(f_I(x)|\mathscr{B}_{SI}). \quad \text{a. s. } \mathscr{P}^S.\end{aligned}$$

引理 5.9 若条件 B 成立，则 $\mathscr{B}_{SI}$ 是关于 $\mathscr{P}^{\mathrm{I}}$ 的充分 σ-域，且在给定 $\mathscr{B}_{SI}$ 条件下，$\mathscr{B}_S$ 与 $\mathscr{B}_I$ 独立.

证明 由条件 B, 存在关于 $\mathscr{B}_S$ 的正则条件概率 $P(B|S)$, 且对 $\mathscr{G}$ 不变，即 $P(gB|\underline{g}S)=P(B|S)$. 如果 $f_I(x)$ 不变,

$$\begin{aligned}\int f_I(y)P(dy|S)&=\int_{g\mathscr{X}} f_I(gy)P(dgy|S)\\&=\int_{\mathscr{X}} f_I(y)P(dgy|S)=\int_{\mathscr{X}} f_I(y)P(dy|\underline{g}^{-1}S).\end{aligned}$$

这就证明了 $f_{SI}(x)\triangleq\int f_I(y)P(dy|S)$ 为 $\mathscr{B}_{SI}$ 可测函数，且由条件期望与条件概率函数的关系知

$$f_{SI}(x)=E(f_I|\mathscr{B}_S). \quad \text{a. s. } \mathscr{P}^S.$$

利用前一引理中 i)与 iii)的等价性，此引理即得证.

引理 5.7 与 5.9 的结论均为 $\mathscr{B}_{SI}$ 对 $\mathscr{P}^{\mathrm{I}}$ 是充分的，而定理 5.3 的条件使

$V=U(S)$的生成域即是$\mathscr{B}_{SI}$，从而V是不变充分统计量．从而定理5.3得证．

附带指出：易见可由引理5.4或定理5.1的条件来保证条件A、B成立．

例5.1 设$X\sim N(\mu, A)$，$A>0$，$\mu\in R^k$皆未知，$X_1, \cdots, X_n$是n次随机独立从X的抽样．众所周知，$\bar{x}=\sum\limits_{i=1}^n x_i/n$，$S^2=\sum\limits_{i=1}^n (x_i-\bar{x})(x_i-\bar{x})^\tau$为$(\mu, A)$的充分统计量．在仿射群

$$\mathscr{G}=\left(\begin{pmatrix}1 & 0\\ e & C\end{pmatrix}: e\in R^k,\ \begin{matrix}C\text{为下三角阵，}\\ \text{对角元素为正}\end{matrix}\right)$$

下，引起参数空间的变换群$\bar{\mathscr{G}}=\mathscr{G}$，行动空间变换群采用$\mathscr{G}^*=\mathscr{G}$．如本章所述，在充分统计量的值域引起变换群$\underline{\mathscr{G}}=\mathscr{G}$，对$\forall g\in\mathscr{G}$，$\underline{\mathscr{G}}(\bar{x}, S^2)=(C\bar{x}+e, CS^2C^\tau)$，其它$\mathscr{G}$，$\bar{\mathscr{G}}$，$\mathscr{G}^*$在其相应变换下的作用与§2中4所述相同．由于$(\mathscr{X}, \mathscr{B}_{\mathscr{X}})$及充分统计量$T=(\bar{x}, S^2)$的值域空间$(\mathscr{T}, \mathscr{B}_{\mathscr{T}})$皆为有限维欧氏空间，引理5.5的条件成立，因此条件B成立．

令S表示下三角阵，对角元素为正，且$S^1S^\tau=S^2$，对任意不变统计量$f(x_1, \cdots, x_n)$，取$g=\begin{pmatrix}1 & 0\\ \bar{x} & S\end{pmatrix}$，

$$f(x_1, \cdots, x_n)=f(g^{-1}(x_1, \cdots, x_n))=f(S^{-1}(x_1-\bar{x}), \cdots, S^{-1}(x_n-\bar{x})).$$

故$(S^{-1}(x_1-\bar{x}), \cdots, S^{-1}(x_n-\bar{x}))$是最大不变量．同样可论证只依赖于充分统计量$(\bar{x}, S^2)$的不变统计量$f(\bar{x}, S^2)$只能是常数(因为对上述$g$，$f(g(\bar{x}, S))=f(0, I)$)．因此$\beta_{SI}=\beta_S\cap\beta_I$仅含有两个元素$\{\emptyset, \mathscr{X}\}$，是最简单的$\sigma$域，由引理5.9可知$\beta_{SI}$是不变充分$\sigma$域，并且$\bar{x}$、$S^2$与$(S^{-1}(x_1-\bar{x}), \cdots, S^{-1}(x_n-\bar{x}))$独立．

§6* 扩充的minimax原则及Hunt-Stein定理

这一节我们进一步推广§3与§4的结果，讨论极小风险估计(最优同变估计)的minimax性质．一方面对变换群进行推广；另一方面对minimax原则进行扩充．现在仍考虑在某种变换群的不变判决问题．$\mathscr{G}$、$\bar{\mathscr{G}}$、$\mathscr{G}^*$分别是对样本空间$(\mathscr{X}, \mathscr{B}_{\mathscr{X}})$、参数空间$(\Theta, \mathscr{B}_\Theta)$、行动空间$(D, \mathscr{B}_D)$的可测变换群，且$\mathscr{G}$与$\bar{\mathscr{G}}$同态；$\mathscr{G}$与$\mathscr{G}^*$同态．$(\mathscr{X}, \mathscr{B}_{\mathscr{X}})$上有分布族$\mathscr{P}=\{P_\theta, \theta\in\Theta\}$，对$\forall g\in\mathscr{G}$，$gx$的分布仍属于$\mathscr{P}$．由$g$对应的$\bar{g}$有$\bar{g}\Theta=\Theta$，但在$\Theta$可产生不同的轨道$\Theta_S$，这些轨道彼此不交，$\bigcup\limits_S\Theta_S=\Theta$，且对$\forall\bar{g}\in\bar{\mathscr{G}}$及$S$，有$\bar{g}\Theta_S=\Theta_S$．在§1中已证明了在不变统计判决问题中，同变估计在每个轨道上有一个常

数风险.

一个估计 δ_1 在扩充的 minimax 意义下不次于另一估计 δ_2, 如果

$$\sup_{\theta\in\Theta_s} R(\theta,\ \delta_1)\leqslant \sup_{\theta\in\Theta_s} R(\theta,\ \delta_2)$$

对所有轨道 Θ_S 成立.

一个估计类 $\mathscr{D}^*$ 称之为扩充的 minimax 意义下的基本完全类,如果对任意估计 δ, 皆存在 $\delta^*\in\mathscr{D}^*$, 使得在扩充的 minimax 意义下 δ^* 不次于 δ.

在本节要指出,同变随机化估计类将构成上述意义的基本完全类. 对于假设检验, Hunt-Stein[9] 首先处理了这一问题, 对于估计问题, Kiefer[11] 在 1957 年做了较好的工作, 这里介绍 Wesler[18] 的工作, 他推广了 Hunt-Stein 的结果. 但我们对条件和证明都作了改进. 本节需要的某些拓扑学等方面知识, 读者可参阅 [7]、[10]、[13].

基本假定:除开使判决问题在群 $\mathscr{G}$ 下保持不变的假定外,还假设:

假设 A_1: 样本空间$(\mathscr{X},\mathscr{B}_{\mathscr{X}})$是可分的,分布族 $\mathscr{P}=\{P_\theta,\theta\in\Theta\}$ 与$(\mathscr{X},\mathscr{B}_{\mathscr{X}})$上的某个 σ 有限测度 μ 等价,即 $\mathscr{P}\equiv\mu$. 这假设是常被满足的(如欧氏空间上有公共支撑的受控分布族).

假设 A_2: 行动空间 D 是可分的距离空间(指存在可数个处处稠密的子集), 具有局部紧拓扑结构, $\mathscr{B}_D$ 由 D 上紧子集所生的最小 σ 域[$(D,\ \mathscr{B}_D)$是欧氏空间满足 A_2].

假设 A_3: 群 $\mathscr{G}$ 有由它的子集所构成的 σ 域 $\mathscr{B}_g$, 在$(\mathscr{G},\ \mathscr{B}_g)$上存在一串概率测度 $\{\nu_n\}$, 使得对于 $\forall g\in\mathscr{G}$ 及 $G\in\mathscr{B}_g$, 皆有

$$\lim_{n\to\infty}\{\nu_n(Gg)-\nu_n(G)\}=0.$$

称为*存在渐近不变测度序列*, 而且 $\mathscr{G}^*$ 作用于行动空间是连续变换, 亦即对于 $\forall g^*\in\mathscr{G}^*$ 及任意 D 的开子集 U、g^*U 也是开集.

(若 $\mathscr{G}$ 是局部紧拓扑群,则假设前半成立).

假设 A_4: 损失函数满足: 对于任意的 $\theta\in\Theta$, 则有 h_0, 当 $h\geqslant h_0$ 使 $\{a:L(\theta,\ a)\leqslant h\}$ 是紧集, 一般地, $\{a:L(\theta,\ a)\leqslant h\}$ 关于 h 是几乎处处(对 Lebesgue 测度)为紧集.

假设 A_5: $\mathscr{G}$ 是局部紧拓扑群(例如线性变换群), $\mathscr{B}_g$ 由所有紧子集所生成.

定理6.1 设 $\delta(\cdot|x)$是 θ 的任一随机化估计, 对任意固定的 $A\in\mathscr{B}_D$, $g\in\mathscr{G}$, $\delta(g^*A|gx)$是关于 $\mathscr{B}_g\times\mathscr{B}_{\mathscr{X}}$ 可测. 则在假设 $A_1\sim A_4$ 之下,存在几乎同变估计 $\delta^*(\cdot|x)$在扩充的 minimax 意义下不次于 $\delta(\cdot|x)$. 若进一步假设 A_5 成立,则存在同变估计 $\delta^*(\cdot|x)$在扩充的 minimax 意义下不次于 $\delta(\cdot|x)$.

证明 分为两种情况来证:

情况 I: 若 $\mathscr{G}$ 是紧的拓扑群, 例如有限群, 由此在 $\mathscr{G}$ 上存在一个右不变 Haar 概率测度 ν, 即对于 $\forall G\in\mathscr{B}_g$, $g\in\mathscr{G}$, $\nu(Gg)=\nu(G)$. 现对 $A\in\mathscr{B}_D$, $x\in\mathscr{X}$, 作

$$\delta^*(A|x)=\int_g \delta(g^*A|gx)d\nu(g). \tag{6.1}$$

不难看到, 如此定义的 $\delta^*(\cdot|x)$ 仍然构成随机化估计, 即对于固定的 x, 它是 $\mathscr{B}_D$ 上的概率测度, 对固定 $A\in\mathscr{B}_D$, 由于 $\delta(g^*A|gx)$ 对于 $\mathscr{B}_g\times\mathscr{B}_{\mathscr{X}}$ 可测, 它是 $\mathscr{B}_{\mathscr{X}}$ 上的可测函数. 而且 δ^* 同变, 这是因为对于 $\forall g\in\mathscr{G}$, $A\in\mathscr{B}_D$, 有

$$\begin{aligned}\delta^*(g^*A|gx)&=\int_g \delta(g_1^*g^*A|g_1gx)d\nu(g_1)\\&=\int_g \delta(g_2^*A|g_2x)d\nu(g_2g^{-1})\\&=\int_g \delta(g_2^*A|g_2x)d\nu(g_2)=\delta(A|x).\end{aligned}$$

现在任取 $\theta_0\in\Theta$, Θ_{s_0} 为含 θ_0 的轨道, 则

$$\begin{aligned}R(\theta_0,\ \delta^*)&=\int_{\mathscr{X}}\int_D L(\theta_0,\ a)\delta^*(da|x)P_{\theta_0}(dx)\\&=\int_{\mathscr{X}}\int_{\mathscr{G}}\int_D L(\theta_0,\ a)\delta(dg^*a|gx)d\nu(g)P_{\theta_0}(dx)\\&=\int_{\mathscr{G}}\int_{\mathscr{X}}\int_D L(\bar{g}\theta_0,\ a)\delta(da|x)P_{\bar{g}\theta_0}(dx)d\nu(g)\\&\leqslant\sup_{\bar{g}\in\bar{\mathscr{G}}} R(\bar{g}\theta_0,\ \delta)=\sup_{\theta_0\in\Theta_{s_0}} R(\theta_0,\ \delta).\end{aligned} \tag{6.2}$$

这就在 $\mathscr{G}$ 为紧拓扑群的情况下证明了定理.

情况 II: 一般情况, $\mathscr{G}$ 不是紧拓扑群, 不存在右不变 Haar 概率测度.

令 $\mathscr{K}$ 是 D 上所有紧子集的类, 由假设 A_3, 在($\mathscr{G}$, $\mathscr{B}_g$)上存在一串渐近不变的概率测度 $\{\nu_n\}$, 对定理给出的 $\delta(\cdot|x)$, 作

$$\delta_n(A|x)=\int\delta(g^*A|gx)d\nu_n(g),\ \forall A\in\mathscr{K},\ x\in\mathscr{X},$$

根据弱紧定理(参见[12]附录 4), 对每个 $A\in\mathscr{K}$, δ_n 在弱紧意义下有聚点存在, 即存在子序列 $\{\delta_{n_i}(A|x)\}$ 及 $0\leqslant\delta_0(A|x)\leqslant1$ (固定 A 时它为 $\mathscr{B}_{\mathscr{X}}$ 上的可测), 使得对于任意 μ 可积函数 f, 有下式成立:

$$\lim_{i\to\infty}\int\delta_{n_i}(A|x)f(x)d\mu=\int\delta_0(A|x)f(x)d\mu, \tag{6.3}$$

亦即所有 $0\leqslant\delta(A|x)\leqslant1$ 的泛函空间 Ψ_A 在弱紧意义下是紧空间. 再由 Tychonoff 紧空间的定理知(参见[10]): 由每个 Ψ_A 的紧性得到乘积空间 $\prod_{A\in K}\Psi_A$ 在乘积拓扑意义下也是紧空间. 由此知, 对于 $\delta_n(A|x)$, 存在 n 的子

列 $\{n_j\}$ 与 A 无关及 $\delta_0(\cdot|x)$，使得对于任何固定的 $A\in\mathscr{X}$，$\delta_0(A|x)$ 为 $\mathscr{B}_{\mathfrak{X}}$ 可测，固定 x 它是 $\mathscr{X}$ 上集合的函数，并对任一 μ 可积函数 f，有

$$\lim_{j\to\infty}\int\delta_{n_j}(A|x)f(x)d\mu=\int\delta_0(A|x)f(x)d\mu. \tag{6.4}$$

由于 $\delta_n(\cdot|x)$ 对固定的 x 是概率测度，从而可推得下列等式或不等式 a. s. μ 成立：

a) 若 $A_1\subset A_2$，A_1，$A_2\in\mathscr{X}$，则 $\delta_0(A_1|x)\leqslant\delta_0(A_2|x)$；

b) $0\leqslant\delta_0(A|x)\leqslant1$，当 $A\in\mathscr{X}$；

c) 当 A_1、$A_2\in\mathscr{X}$，$A_1\cap A_2=\emptyset$，则

$$\delta_0(A_1\cup A_2|x)=\delta_0(A_1|x)+\delta_0(A_2|x).$$

a) 与 b) 的结论是显然的，现在证明 c)，因为 μ 是 σ 有限测度，取 $f(x)>0$, a. s. $[\mu]$，且 μ 可积，定义

$$g(x)=f(x)\{\delta_0(A\cup A_2|x)-\delta_0(A_1|x)-\delta_0(A_2|x)\}.$$

显然 $g(x)$ 也关于 μ 可积，对任意的 n，有

$$0=\int g(x)[\delta_{n_j}(A\cup A_2|x)-\delta_{n_j}(A_1|x)-\delta_{n_j}(A_2|x)]d\mu.$$

从而得出

$$0=\int g(x)[\delta_0(A_1\cup A_2|x)-\delta_0(A_1|x)-\delta_0(A_2|x)]d\mu$$

$$=\int f(x)[\delta_0(A_1\cup A_2|x)-\delta_0(A_1|x)-\delta_0(A_2|x)]^2d\mu.$$

由于上式被积函数是非负的及 $f(x)>0$, a. s. μ，即得 c)的结论.

下面根据 $\delta_0(\cdot|x)$ 来构造一个适当的判决函数 δ^*：由于 D 是可分距离空间，故存在可数子集在 D 处处稠密. 用 R 表示所有以此稠密子集的点为中心、有理数为半径的闭球的有限和集. 从而 D 的开子类可表为属于 R 的可列个集合内核的和集. 由于 R 只有可数个集合，故除去一个 μ 零测度集 N 外，在 R 上满足 a)、b)、c)三条性质，令 D 的开子集类为 $\mathscr{U}$，对 $\forall U\in\mathscr{U}$，$x\bar{\in}N$，定义

$$\delta^*(U|x)=\sup_{A\in R,\,A\subset U}\delta_0(A|x). \tag{6.5}$$

显然对于固定的 U，$\delta^*(U|x)$ 是 $B_{\mathfrak{X}}$ 可测函数，又由于 $\delta^*(U|x)$ 有半可加性，故可对 D 的任一子集 W，定义

$$\delta^*(W|x)=\inf_{W\subset U\in u}\delta^*(U|x). \tag{6.6}$$

无疑，对固定的 W，$\delta^*(W|x)$ 也是 $B_{\mathfrak{X}}$ 可测函数，现在分几步证明此 $\delta^*(\cdot|x)$ 的几乎同变、不次于 δ.

(一)除去一个 μ 零测度集外，$\delta^*(\cdot|x)$ 是 $(D,\mathscr{B}_D)$ 上的概率测度.

i) 当 $x\overline{\in}N$ 时，$\delta^*(U|x)$在 D 的所有开子集类 $\mathscr{U}$ 上构成测度。

设 $U_i\in\mathscr{U}(i=1, 2, \cdots)$当 $x\overline{\in}N$ 时，由 δ^* 的定义就可得

1) 若 $U_1\subset U_2\Rightarrow\delta^*(U_1|x)\leqslant\delta^*(U_2|x)$;

2) $\delta^*(\phi|x)=0$;

3) 若 $U_1\cap U_2=\emptyset\Rightarrow\delta^*(U_1\cup U_2|x)=\delta^*(U_1|x)+\delta^*(U_2|x)$;

4) 若 $U_i\cap U_j=\emptyset(i\neq j)$，则 $\delta^*\left(\bigcup\limits_{i=1}^{\infty}U_i|x\right)=\sum\limits_{i=1}^{\infty}\delta^*(U_i|x)$。

1)与 2)是显然的，现在就 3)与 4)给出证明:

对于 $U_1\cup U_2$，由定义 (6.6) 知，对于 $x\overline{\in}N$，任给 $\varepsilon>0$，存在 $A\in R$，$A\subset U_1\cup U_2$，使得 $\delta_0(A|x)+\varepsilon\geqslant\delta^*(U_1\cup U_2|x)$，对 A 的每个点 a，存在有理数 r_a，使开球 $\{x:\|x-a\|<r_a\}$ 或含于 U_1 或含于 U_2，所以开球 $U_a=\{d:\|d-a\|<r_a/2\}$ 及其闭包 $\overline{U}_a$ 仍含于 U_1 或含于 U_2。这样的开球 U_a 把 A 覆盖了。由于 A 是紧集，根据有限覆盖定理，存在有限个点 $a_1, \cdots, a_k\in A$，使得

$$A\subset\bigcup_{i=1}^{k}U_{a_i}\subset\bigcup_{i=1}^{k}\overline{U}_{a_i}\subset U_1\cup U_2,$$

$$V_1\triangleq\bigcup_{a_j\in U_1}\overline{U}_{a_j}\subset U_1,\ V_2\triangleq\bigcup_{a_i\in U_2}\overline{U}_{a_i}\subset U_2.$$

注意：$V_i\in R(i=1, 2)$，$V_1\cup V_2\supset A$，所以

$$\begin{aligned}\delta^*(U_1|x)+\delta^*(U_2|x)&\geqslant\delta^*(V_1|x)+\delta^*(V_2|x)\\&=\delta^*(V_1\cup V_2|x)\geqslant\delta^*(A|x)\geqslant\delta^*(U_1\cup U_2)-\varepsilon.\end{aligned}$$

根据 ε 的任意性可知

$$\delta^*(U_1|x)+\delta^*(U_2|x)\geqslant\delta^*(U_1\cup U_2|x).$$

由定义(6.6)，又显然可得

$$\delta^*(U_1|x)+\delta^*(U_2|x)\leqslant\delta^*(U_1\cup U_2|x).$$

所以 3) 得证. 同时用归纳法就可证明当 $U_i\in\mathscr{U}(i=1, 2, \cdots)$ 时为彼此不交的开集，则对任意的 n，有

$$\delta^*\left(\bigcup_{i=1}^{n}U_i|x\right)=\sum_{i=1}^{n}\delta^*(U_i|x)\leqslant\delta^*\left(\bigcup_{i=1}^{\infty}U_i|x\right)\quad(\text{当 } x\overline{\in}N).$$

令 $n\to\infty$，则有

$$\sum_{i=1}^{\infty}\delta^*(U_i|x)\leqslant\delta^*\left(\bigcup_{i=1}^{\infty}U_i|x\right)\quad(x\overline{\in}N).$$

再用证明 3)同样的办法也可证明

$$\sum_{i=1}^{\infty}\delta^*(U_i|x)\geqslant\delta^*\left(\bigcup_{i=1}^{\infty}U_i|x\right)\quad(x\overline{\in}N).$$

至此 4)得证.

ii) 当 $x\overline{\in}N$ 时，$\delta^*(\cdot|x)$是$(D, \mathscr{B}_D)$上的测度.

根据 $\delta^*(\cdot|x)$的定义及 i)，它是外测度. 如同在实变函数中建立 L 测

度那样，一个 D 的子集 E 称为可测，若对 D 的任意子集 A, 有下式成立

$$\delta^*(A|x)\geqslant\delta^*(AE|x)+\delta^*(AE^c|x). \tag{6.7}$$

这里 $E^c=D-E$. 由于 D 是可分距离空间，从外测度的性质易知，D 的子集 E 为可测的充要条件是：(6.7)式仅对每个开集 U 代 A 成立. 事实上，必要性是显然的，只要证充分性：对 D 的任意子集 A, 有 $U\in\mathscr{U}$ 使 $A\subset U$, 由假设知

$$\delta^*(U|x)\geqslant\delta^*(UE|x)+\delta^*(UE^c|x)\geqslant\delta^*(AE|x)+\delta^*(AE^c|x).$$

上式右端对 $\{U:A\subset U,\ U\in\mathscr{U}\}$ 取下确界，由定义(6.6)即知应有(6.7)式成立.

对于 D 的任意紧子集 E, 要证 E 可测：对于开集 U, UE^c 为开集，由(6.5)式知：存在闭集 $F_1\subset CE^c$, 使

$$\delta^*(F_1|x)\geqslant\delta^*(UE^c|x)-\varepsilon.$$

又 UF_1^c 也为开集，从而存在 $F_2\subset UF_1^c$, 使

$$\delta^*(F_2|x)\geqslant\delta^*(UF_1^c|x)-\varepsilon.$$

显然 $F_1\cup F_2\subset U$, $F_1\cap F_2=\emptyset$. $UF_1^c\supset UE$, 故

$$\begin{aligned}\delta^*(U|x)&\geqslant\delta^*(F_1\cup F_2|x)=\delta^*(F_1|x)+\delta^*(F_2|x)\\&\geqslant\delta^*(UE^c|x)+\delta^*(UE|x)-2\varepsilon.\end{aligned}$$

由于 ε 的任意性，上式即给出 E 为可测，而 D 的所有可测子集也构成 σ 域，$\mathscr{B}_D$ 是由 D 的所有紧子集生成的最小 σ 域，因而 $\mathscr{B}_D$ 的元皆为可测集. 这就说明了当 $x\overline{\in}N$ 时，$\delta^*(\cdot|x)$ 是 $\mathscr{B}_D$ 上的测度.

iii) 存在 μ 零测度集 N_1, 当 $x\overline{\in}N\cup N_1$ 时，$\delta^*(\cdot|x)$ 是 $(D,\mathscr{B}_D)$ 上的概率测度.

首先指出：对每个紧集 $A\in\mathscr{K}$, $x\overline{\in}N$, 令开集 $U\supset A$, 如同 i)中证明的那样，存在 $\overline{A}_{a_1},\cdots,\overline{A}_{a_k}$(分别是以 $a_i(i=1,\cdots,k)$ 为球心的属于 R 的闭球). $\bigcup\limits_{i=1}^k\overline{A}_{a_i}\in R$, 且 $A\subset\bigcup\limits_{i=1}^k\overline{A}_{a_i}\subset U$, 因此

$$\delta_0(A|x)\leqslant\delta_0\Big(\bigcup_{i=1}^k\overline{A}_{a_i}|x\Big)\leqslant\delta^*(U|x).$$

由 δ^* 的定义(6.6)即知

$$\delta_0(A|x)\leqslant\delta^*(A|x). \tag{6.8}$$

不失一般性，现设对于定理给出的 $\delta(\cdot|x)$, 存在 Θ 在 $\mathscr{G}$ 变换下的某一轨道 Θ_S 及 $0<m_S<\infty$, 使之 $R(\theta,\ \delta)\leqslant m_S$, $\forall\theta\in\Theta$(否则，定理自然成立). 现取 $\theta_0\in\Theta_S$ 及适当小的 $\varepsilon>0$, 定义 $A=\{a:L(\theta_0,\ a)\leqslant m_S/\varepsilon\}$, 根据假设 A_3 知：存在$(\mathscr{G},\ \mathscr{B}_g)$上概率测度 ν_n, 与导出(6.2)式同样的步骤可导出 $R(\theta_0,\ \delta_n)\leqslant m_S$, 从而

$$m_S \geqslant R(\theta_0, \delta_n) \geqslant \iint_{D-A} \delta_n(da|x) L(\theta_0, a) P_{\theta_0}(dx)$$

$$\geqslant m_S \varepsilon^{-1} \int_{\mathscr{X}} \int_{D-A} \delta_n(da|x) P_{\theta_0}(dx)$$

$$\geqslant m_S \varepsilon^{-1} \int \delta_n(D-A|x) P_{\theta_0}(dx),$$

即 $\int_{\mathscr{X}} \delta_n(D-A|x) P_{\theta_0}(dx) < \varepsilon$，从而对一切 n，有

$$\int_{\mathscr{X}} \delta_n(A|x) P_{\theta_0}(dx) > 1-\varepsilon.$$

再由 δ_0 的定义，当 ε 充分小时，$A \in \mathscr{K}$ 及(6.8)式，得

$$\int \delta^*(D|x) P_{\theta_0}(dx) \geqslant \int \delta^*(A|x) P_{\theta_0}(dx) \geqslant \int \delta_0(A|x) P_{\theta_0}(dx) \geqslant 1-\varepsilon.$$

由 ε 的任意性，$0 \leqslant \delta^* \leqslant 1$ 以及 $\mathscr{P} \equiv \mu$，故 $\delta^*(D|x)=1$, a. s. μ. 令 $N_1 = \{x:\delta^*(D|x) \neq 1\}$，则当 $x \in N \cup N_1$ 时，$\delta^*(\cdot|x)$ 为概率测度.

(二) $\delta^*(\cdot|x)$ 是几乎同变的.

i) 对于每个 $(\mathscr{G}, \mathscr{B}_g)$ 上的有界可测函数 $\psi(g)$，有

$$\lim_{n\to\infty} \int_{\mathscr{G}} \psi(g')\{d\nu_n(g'g^{-1}) - d\nu_n(g')\} = 0.$$

此时不妨设 $|\psi| \leqslant 1$，把 $[-1, 1]$ 等分为 $2M$ 个小区间，由积分定义知

$$\Big| \sum_{k=-M}^{M} \frac{k}{M} \Big[\nu_n\Big(\Big(\frac{k}{M} \leqslant \psi(g') < \frac{k+1}{M}\Big) g^{-1}\Big) - \nu_n\Big(\frac{k}{M} \leqslant \psi(g') < \frac{k+1}{M}\Big)\Big]$$

$$- \int \psi(g')[d\nu_n(g'g^{-1}) - d\nu_n(g')] \Big| \leqslant \frac{2}{M}.$$

令 $n \to \infty$，则上式左端第一项趋于 0，再令 $M \to \infty$，则证明了 i)的结论.

ii) 对每个 $g \in \mathscr{G}$ 及紧集 A，有

$$\delta_0(g^* A|gx) = \delta_0(A|x), \text{ a. s. } \mu. \tag{6.9}$$

令 $f(x)$ 是任意关于 μ 的可积函数. 由上面(二)的 i)，有

$$\int [\delta_0(g^*A|gx) - \delta_0(A|x)] f(x) d\mu$$

$$= \lim_{n\to\infty} \iint \delta(g'^* A|g'x) f(x) d\mu [d\nu_n(g'g^{-1}) - d\nu_n(g')] = 0.$$

其中要注意 $\psi(g') = \int \delta(g'^*A|g'x) f(x) d\mu$ 是有界可测函数，取 $f(x) = [\delta_0(g^*A|gx) - \delta_0(A|x)] g(x)$，$g(x) > 0$ 可积，即如证明 $\delta_n(\cdot|x)$ 的性质 c)那样可证得 ii)的结论.

iii) 对每个 $g \in \mathscr{G}$ 及每个紧集 A，有

$$\delta^*(g^*A|gx) = \delta^*(A|x), \text{ a. s. } \mu. \tag{6.10}$$

令 U 是一个开集，$A\subset \bar{U}$，按 $\delta^*(A|x)$ 的定义(6.5)，存在 $F_n\in\bar{\mathscr{R}}$，且 $F_n\subset U(n=1, 2, \cdots)$，有

$$\lim_{n\to\infty}\delta_0(F_n|x)=\delta^*(U|x),\quad \lim_{n\to\infty}\delta_0(g^*F_n|gx)=\delta^*(g^*U|gx).$$

由 ii)的结论即知 $\delta^*(U|x)$ 与 $\delta^*(g^*U|gx)$. a. s. μ 相等，这样可见

$$\delta^*(A|x)\leqslant\delta^*(U|x)=\delta^*(gU|gx),\ \text{a. s. }\mu.$$

$$\delta^*(g^*A|gx)\leqslant\delta^*(g^*U|gx)=\delta^*(U|x),\ \text{a. s. }\mu.$$

在上式两边对 U 取 inf，即可得 iii)的结论.

iv) 几乎同变.

令 $\mathscr{A}$ 为使 $\delta^*(g^*B|gx)=\delta^*(B|x)$, a. s. μ. 的所有 $B\in\mathscr{B}_D$ 的集合类，易证这一类构成单调类，由 $\mathscr{A}\supset\mathscr{K}$ 知 $\mathscr{B}_D=\mathscr{A}(\mathscr{K})\subset\mathscr{A}$，此处 $\mathscr{A}(\mathscr{K})$ 是集合类 $\mathscr{K}$ 生成的最小单调类，由测度论的知识知：它就等于由 $\mathscr{K}$ 生成的 σ 域，即 $\mathscr{B}_D$. 说明 δ^* 几乎同变.

(三) δ^* 不次于 δ.

因为

$$\begin{aligned}R(\theta, \delta^*)&=\int_0^\infty\int_{\mathscr{X}}\delta^*\{[a:L(\theta, a)>h]|x\}p(x|\theta)d\mu dh\\&=\int_0^\infty\int_{\mathscr{X}}(1-\delta^*\{[a:L(\theta, a)\leqslant h]|x\})p(x|\theta)d\mu dh\\&=\lim_{H\to\infty}\int_0^H\int_{\mathscr{X}}(1-\delta^*\{[a:L(\theta, a)\leqslant h]|x\})p(x|\theta)d\mu dh\\&=\lim_{H\to\infty}\left(H-\int_0^H\int_{\mathscr{X}}\delta^*\{[a:L(\theta, a)\leqslant h]|x\}\right)p(x|\theta)d\mu dh.\end{aligned}\quad(6.11)$$

根据假设 A_4，$\{a:L(\theta, a)\leqslant h\}$ 对 h 而言是几乎(Lebesgue 测度)为紧集. 对紧集 F，由 $\delta^*(F|x)\geqslant\delta_0(F|x)$，而且存在子序列 δ_{n_k}，使得

$$\lim_{k\to\infty}\int\delta_{n_k}(F|x)p(x|\theta)d\mu=\int\delta_0(F|x)p(x|\theta)d\mu.$$

由(6.11)及上述两点，则得

$$\begin{aligned}R(\theta, \delta^*)&\leqslant\lim_{H\to\infty}\left(H-\int_0^H\int_{\mathscr{X}}\delta_0\{[a:L(\theta, a)\leqslant h]|x\}\right)p(x|\theta)d\mu dh\\&=\lim_{H\to\infty}\left(H-\int_0^H\lim_{k\to\infty}\int\delta_{n_k}\{[a:L(\theta, a)\leqslant h]|x\}\right)p(x|\theta)d\mu dh\\&=\lim_{H\to\infty}\lim_{k\to\infty}\left(H-\int_0^H\int\delta_{n_k}\{[a:L(\theta, a)\leqslant h]|x\}\right)p(x|\theta)d\mu dh\\&=\lim_{H\to\infty}\lim_{k\to\infty}\int_0^H\int\delta_{n_k}\{[a:L(\theta, a)>h]|x\}p(x|\theta)d\mu dh\\&\leqslant\lim_{k\to\infty}R(\theta, \delta_{n_k})\leqslant\sup_{\theta\in\Theta_s}R(\theta, \delta)\quad(\text{当 }\theta\in\Theta_s\text{ 时}).\end{aligned}$$

注意：当 $\{a:L(\theta, a)\leqslant h\}$ 不是紧集时，$\delta_0\{[a:L(\theta, a)\leqslant h]|x\}$ 没有定义，我

们可定义它为 1、根据假设 A_4，并不影响上述不等式的成立.

至此定理第一部分得证.

至于定理第二部分：由于 D 为可分距离空间，如同证明(一)i)的办法，对每个紧集 A 存在以 a_i 为球心的闭球 $\bar{A}_{a_i}\in R(i=1, \cdots, k)$，$\bigcup_{i=1}^{k}\bar{A}_{a_i}\supset A$，因为 $\bigcup_{i=1}^{k}\bar{A}_{a_i}\in R$，所以 A 可表为 R 中一个序列元的下界，因此 $\mathscr{B}(\mathscr{X})=\mathscr{B}(R)$. 这说明 $\mathscr{B}_D$ 是可分的. 再根据假设 A_5，在$(\mathscr{G}, \mathscr{B}_g)$上存在右 Haar 测度 ν. 对任意的 $G\in\mathscr{B}_g$，$\forall g\in\mathscr{G}$，$\nu(Gg)=\nu(G)$，如果能够证明 $\delta^*(g^*A|gx)$对固定的 $A\in\mathscr{B}_D$ 为 $\mathscr{B}_g\times\mathscr{B}_{\mathscr{X}}$ 可测，则由 § 5 引理 5 的证明思路可知：存在一个μ零测度集 N 及同变估计 $\delta_I^*(\cdot|x)$与 $\delta(\cdot|x)$等价，即

$$\delta_I^*(A|x)=\delta^*(A|x), \text{ 当 } x\overline{\in} N,\ A\in\mathscr{B}_D.$$

这样，定理的第二部分得证.

例 6.1 在 § 2 问题 4 所举的估计多元正态分布协方差阵 A 的例子中，已求出了在仿射群下最优同变估计. 由于损失函数 $t_r(A^{-1}D-I)^2$ 是严格凸函数. 群 $\mathscr{G}$ 是线性变换群. 由 § 5 的讨论知：最优同变估计是唯一的. 加上此问题仅有唯一的轨道，以及样本空间$(\mathscr{X}, \mathscr{B}_{\mathscr{X}})$与判决空间都是欧氏空间结构，群 $\mathscr{G}$ 是局部紧拓扑群，因此定理的条件都满足，故 A 的最优同变估计是 minimax 估计.

问题与习题

1. 设 $x_1, x_2, \cdots, x_n$ 是 iid，有极值分布，其概率密度为

$$f(x, \alpha)=\exp(-(x-\alpha)-e^{-(x-\alpha)}) \quad (-\infty<x<\infty,\ -\infty<\alpha<\infty).$$

$\mathscr{G}$ 为转移群，损失函数为 $L(\alpha, \hat{\alpha})=(\alpha-\hat{\alpha})^2$. 设 $\hat{\alpha}(x_1, \cdots, x_n)$是基于极小充分统计量 Σx_i, Σe^{-x_i}的同变估计. 求极小风险的最优同变估计.

2. 考虑转移型位置参数的 Pareto 分布，其密度函数是

$$f(x, \alpha)=\frac{3}{\alpha}\left(\frac{\alpha}{x}\right)^2 \quad (\alpha\leqslant x<\infty,\ 0<\alpha<\infty).$$

对于平方损失，找最优同变估计.

3. 如果行动空间$(D, \mathscr{B}_D)$是可分的(指 $\mathscr{B}_D$ 可分)，试证定理 1.1 对几乎同变估计成立.

4. 设 $\{P_\theta: -\infty<\theta<\infty\}$ 是一族 Cauchy 分布. $x_1, \cdots, x_n$ 是 iid，具有 Cauchy 分布.

i) θ 的 Pitman 估计是否存在?

ii) 若对 Cauchy 分布采用截断方法如下:

$$f^{(k)}(x,\theta)=\begin{cases}C\pi^{-1}[1+(x-\theta)^2]^{-1}, & |x-\theta|\leqslant k;\\ 0, & \text{其它}.\end{cases}$$

此处 C 是已知因子(依赖 k), 试求其 Pitman 估计.

5. 令 $x_1, \cdots, x_n$ 为 iid, 服从正态分布 $N(\mu, \sigma^2)$, 损失函数为: 当 $|\mu-d|<\sigma$ 时为 0; 当 $|\mu-d|\geqslant\sigma$ 时为 1, 求 μ 的同变 minimax 估计. 若损失函数为: 当 $|d\sigma^{-1}-1|<1$ 为 0; $|d\sigma^{-1}-1|\geqslant 1$ 时为 1. 求 σ 的同变 minimax 估计.

6. 若 $x_1, \cdots, x_n$ 为 iid 随机向量($n\geqslant p+1$), 具有 p 维非退化正态分布 $N(\theta, A)$, 估计协方差 A:

i) 损失函数为

$$L(A,d)=\frac{(t_rBA-d)^2}{2nt_rBABA+A(t_rBA)^2},$$

此处 B 为已知的对称阵, 现已知 $\theta=0$ 时, $d=t_rBS'/(n+2)$ 为 A 的 minimax 估计. 求证 θ 未知时 $d=t_rBS/(n+1)$ 是 A 的 minimax 估计, 这里

$$S'=\sum_{i=1}^n x_ix_i^\tau,\quad S=\sum_{i=1}^n(x_i-\bar{x})(x_i-\bar{x})^\tau.$$

ii) 若损失函数为 $L(A,D)=t_r(AD-I)^2$, 求 $\theta=0$ 以及 θ 未知时 A^{-1} 的 minimax 估计.

7. 令 F 是实轴上的绝对连续分布, 考虑 $F(x)$ 的估计问题, 损失函数为

$$L(F,\hat{F})=\int|F(x)-\hat{F}(x)|^r dx.$$

此处 r 为正整数. 令 $\mathscr{G}$ 是所有一一对应的单调变换群, 试证明: 若 $X_{(1)}\leqslant\cdots\leqslant X_{(n)}$ 是顺序统计量, 则 F 的 minimax 估计是

$$\hat{F}(x)=(j+1)/(n+2),\ \text{若}\ x_{(j)}\leqslant x\leqslant x_{(j+1)}\quad(j=0,\cdots,n).$$

此处 $x_{(0)}=-\infty$, $x_{(n+1)}=\infty$.

(Aggarwar, Ann. Math. Statist., 26: 450~463(1955)).

8. 令 $x_1, \cdots, x_n$ 为 iid. 分布为 $N(\mu, \sigma^2)$; 损失函数为 $(d\sigma^{-a}-1)^2$ ($a>0$ 已知), 则 $C_0(a)S^a$ 是 σ^a 的 minimax 估计, 其中

$$S^2=\sum_{i=1}^n(x_i-\bar{x})^2,\ C_0(a)=\Gamma\left(\frac{n-1}{2}+a\right)\Big/2^{\frac{a}{2}}\Gamma\left(\frac{n-1}{2}+2a\right).$$

9. 若 $x_1, \cdots, x_n$ 为 iid, 遵从指数分布, 密度为

$$\beta^{-1}\exp[-(x-\theta)/\beta]I_{(x\geqslant\theta)},\ \theta\in R_1,\ \beta>0.$$

试证 $[x_i-x_{(1)}]\Big/\left(\sum_{i=1}^n x_i-nx_{(1)}\right)$ $(i=1,\cdots,n)$ 与 $\left(x_{(1)},\ \sum_{i=1}^n x_i-nx_{(1)}\right)$ 独立.

10. 若 $x_1, \cdots, x_n$ 是 iid, 遵从 $(\theta-\sigma/2, \theta+\sigma/2)$ 区间上的均匀分布, 试

证$(x_i-x_1)/(x_{(n)}-x_{(1)})(i=2, \cdots, n)$与$(x_{(n)}, x_{(1)})$是独立的.

11. 在定理5.3中，若进一步假设S是完备统计量，则V也是不变统计量中的完备充分统计量.

12*. 设n维随机向量$X\sim P((x-\theta 1_n)/\sigma)$，且有密度函数$\sigma^{-n}p((x-\theta 1_n)/\sigma)$ $(-\infty<\theta<\infty, \sigma>0)$. 记$(\hat\theta_1(x), \hat\theta_2(x))$为置信度$1-\alpha$的$\theta$的置信区间，若存在置信度为$1-\alpha$的仿射变换下的最优同变区间估计$(\hat\theta_1^0, \hat\theta_2^0)$，则有

$$\min_{(\hat\theta_1,\hat\theta_2)}\sup_{(\theta,\sigma)} E\left(\frac{\hat\theta_2(x)-\hat\theta_1(x)}{\sigma}\middle|\theta, \sigma\right)=\min_{(\hat\theta_1,\hat\theta_2)\in\mathfrak{D}} E\left(\frac{\hat\theta_2(x)-\hat\theta_1(x)}{\sigma}\middle|\theta, \sigma\right)=E(\hat\theta_2^0-\hat\theta_1^0|0, 1).$$

此处$\mathfrak{D}$表示所有置信度为$1-\alpha$的θ的同变估计类(即$(\hat\theta_1(x), \hat\theta_2(x))$分别皆有同变性质的区间估计).

13. 试证明:

i) 当$x_1, \cdots, x_n$为$iid. \sim N(\mu, \sigma^2)$时，置信度为$1-\alpha$的$\mu$的最优同变置信区间估计为

$$\left(\bar{x}-\sqrt{\frac{n-1}{n}}S_n t_{n-1}\left(\frac{\alpha}{2}\right), \ \bar{x}+\sqrt{\frac{n-1}{n}}S_n t_{n-1}\left(\frac{\alpha}{2}\right)\right).$$

这里$t_{(n-1)}(P)$为$n-1$自由度的t分布的上侧P-分位点，而$S_n^2=\sum_{i=1}^n(x_i-\bar{x})^2$.

ii) 当$x_1, \cdots, x_n$为$iid.$遵从$R(\theta-\sigma/2, \theta+\sigma/2)$时，则置信度$1-\alpha$下$\theta$的最优同变置信区间估计为

$$\left(\frac{x_{(1)}+x_{(n)}}{2}-\frac{1}{2}\alpha^{-\frac{1}{n-1}}(x_{(n)}-x_{(1)}), \ \frac{x_{(1)}+x_{(n)}}{2}+\frac{1}{2}\alpha^{-\frac{1}{n-1}}(x_{(n)}-x_{(1)})\right).$$

iii) 当$x_1, \cdots, x_n$为$iid.$，遵从寿命分布，其密度为

$$\beta^{-1}\exp[-(x-\theta)/\beta]I_{(x>\theta)}.$$

则置信度$1-\alpha$的θ的最优同变置信区间估计为

$$\left(x_{(1)}-\frac{n-1}{n}(\alpha^{-\frac{1}{n-1}}-1)\sum_{i=1}^n(x_i-x_{(1)}), \ x_{(1)}\right).$$

14. 在题12的假定下，对于σ的置信区间有着与题12同样的结论. 试以此求出题13中各分布的σ的最优同变置信区间.

[提示: 先用§4的定理4.1的办法证明θ的区间估计$(d_2-d_1/2, d_2+d_1/2)$在损失函数

$$d/\sigma+cI_{[d_2-d_1/2, d_2+d_1/2]}(\theta). \quad (c>0)$$

之下是最优同变置信区间估计，具有minimax性质，然后看c取何值使这样的最优同变估计具有$1-\alpha$的置信度，如此求出的置信区间即为题12所求.]

15. 对回归模型$Y=X\beta+\varepsilon$, $\beta\in R^k$为未知参数. X为$n\times k(n>k)$矩阵，$X^\tau X>0$, ε为n维随机向量.

i) 若 ε 具有密度 $f(z)$，要估计 β，损失函数为 $\|\beta-d\|^2$. 若令 $\mathscr{G}=\{g_c: c\in R^k\}=\bar{g}=\mathscr{G}^*$ 是向量加法群:

$$g_c y=y+Xc,\ \bar{g}_c\beta=c+\beta,\ g_c^* d=d+c.$$

此变换群使统计判决问题不变. 若存在一个同变估计风险有限. 试证 β 的最优同变估计为

$$\hat{\beta}=\int_{R^k}\beta f(y-X\beta)d\beta \Big/ \int_{R^k} f(y-X\beta)d\beta.$$

并证明此估计是 minimax 估计.

ii) 若知 ε 的分布密度为 $f(z/\sigma)\sigma^{-k}$，$\sigma>0$，估计 β 的损失函数为 $\|\beta-d\|^2/\sigma^2$. 令 y 的变换群为

$$\mathscr{G}=\{g_{c,a},\ c\in R^k,\ a>0\}=\bar{g}=g^*.$$

$$g_{c,a}y=ay+Xc,\ \bar{g}_{c,a}(\beta,\ \sigma)=(a\beta+c,\ a\sigma),\ g_{c,a}^* d=ad+c.$$

在此变换群下，统计判决问题不变. 试证明 β 的最优同变估计为

$$\hat{\beta}=\int_{R^k}\int_0^\infty \beta f\left(\frac{y-X\beta}{\sigma}\right)\sigma^{-k-3}d\sigma\, d\beta \Big/ \int_{R^k}\int_0^\infty f\left(\frac{y-X\beta}{\sigma}\right)\sigma^{-k-3}d\sigma\, d\beta$$

它也是 minimax 估计.

[提示: 如同 § 2 Pitman 估计一样来进行. 至于 minimax 性可采用 § 2 和 § 4 的办法，也可直接验证 Hunt-Stein 定理的条件.]

16. 设随机向量 $X\sim F(x-\theta 1)\ \theta\in R^1$ 未知，$1=(1,\ \cdots,\ 1)^\tau_{p\times 1}$，若 θ 的 UMVUE 存在，则 θ 的 Pitman 估计 $\bar{x}-E_0(\bar{x}\mid x_2-x_1,\cdots,x_p-x_1)$ 是 UMVUE 估计.

[提示: i) UMVUE 基本唯一; ii) Pitman 估计为无偏估计; iii) 有一同变估计与几乎同变估计等价.]

参 考 文 献

[1] D. Blackwell: On a class of probability space. Proc. of the Third Berkeley Symposium On Math. Statist and Prob. I: 1～6 (1956), Berkeley and los Angeles: Univ. of California Press.

[2] D. Blackwell and M. A. Givshick: Theory of Games and Statistical Decision. John Wiley, New York (1954).

[3] 陈桂景、陈希孺:《同变性限制下的充分性原则》, 应用数学学报, 5 (1): 85～93 (1982).

[4] 陈希孺:《位置参数的 minimax 估计》, 数学学报, 14(2): 276～290(1964).

[5] Chen Hsi-Ju(陈希孺): On minimax invariant estimation of scale and location parameters, Scientia Sinica. 13(10) (1964).

[6] M. A. Girshisk, L. G. Sarage: Bayes and minimax estimator for quadrate

loss function. Proc. Second Berkeley Symp. Math Statist. Prob. I. 53～74 (1951).

[7] P. R. Halmos: Measure Theory, D. Van Nostrand Com., Inc. (1950). (有中译本,王建华译,科学出版社)

[8] R. B. Hora, R. J. Buhler, «Fiducial theory and inrariant estimation. Ann. Math. Statist., 37, 643～656(1966).

[9] G. A. Hunt, C. Stein: Most stringent tests of statistical hypotheses, Unpublished, 1945.

[10] J. L. Kelley: Genaral Topology. (1955).

[11] J. Kiefer: Invariance minimax sequential estimation and continuous time processes, Ann. Math. Statist., 28, 573～601(1957).

[12] E. L. Lehmann: Testing Statistical Hypotheses, John Wiley, New York, 1959.

[13] L. Nachbin: The Haar Integral, D. Von Nostrand Priceton, 1965.

[14] I. Olkin and J. B. Sellich: Estimating Covariance in a multivariate normal distribution, Statistical Decision Theory and Related Topics, Edited by S. S. Gupta and D. S. Moore (1977).

[15] E. J. Pitman: The estimation of the location and scale parameters of a continuous population of any given form, Biometrika, 30:391～421(1939).

[16] C. Stein: On Sequence of experiment (abstract). Ann. Math. Statist. 19 117～118(1948).

[17] W. J. Hall, R. A. Wijsman, J. K. Ghosh: The relationship between sufficiency and invariance with applications in sequential analysis Ann. ath. Statist. 36: 575～614(1965).

[18] O. Wesler: Invariance theory and a modified minimax principle. Ann. Math. Statist. 30: 1～20(1959).

第八章　容许估计

§1　容许性概论

在这一章我们仍然假设 $(\mathscr{X}, \mathscr{B}_x, \mathscr{P})$，$\mathscr{P}=\{P_\theta, \theta\in\Theta\}$ 表示样本空间(或概率空间)，此处 $\mathscr{P}$ 表示分布族并且被 $(\mathscr{X}, \mathscr{B}_x)$ 上的 σ 有限测度 μ 所控制，θ 是未知参数. 我们往往假设 Θ 为 k 维欧氏空间 R^k 的 Borel 可测子集，例如连通域. $L(\theta, d)$ 表示损失函数，它是建立在参数空间与行动空间 $(\Theta\times D, \mathscr{B}_{\Theta\times D})$ 的 $\mathscr{B}_{\Theta\times D}$ 可测函数. $\delta(x)$ 表示对 θ 的估计(可以是随机化估计). $\mathscr{D}$ 表示所有估计类，有时称它为判决函数空间，$R(\theta, \delta)$ 表示估计 δ 的风险函数. 我们要在这一章中讨论两个估计优劣的比较，以风险大小来建立判决函数空间的某种偏序.

定义1.1　我们称 $\delta_1\in\mathscr{D}$ 处处不比 $\delta_2\in\mathscr{D}$ 差，如果对于 $\forall\theta\in\Theta$，有

$$R(\theta, \delta_1)\leqslant R(\theta, \delta_2). \tag{1.1}$$

如果进一步假设对于某些 $\theta\in\Theta$，(1.1)式中严格不等号成立，则称 δ_1 优于 δ_2，记为 $\delta_1\prec\delta_2$.

我们称 δ_1 与 δ_2 等效，如果对于 $\forall\theta\in\Theta$，(1.1)式等式成立，记为 $\delta_1\approx\delta_2$.

我们称 $\delta_0\in\mathscr{D}$ 是容许估计，如果不存在估计量 $\delta_1\prec\delta_0$. 也就是说，所有处处不比 δ 差的估计都应与 δ_0 等效.

我们称 δ_0 是几乎容许估计，如果对任何处处不比 δ_0 差的估计 δ，都有

$$L(\theta: R(\theta, \delta)<R(\theta, \delta_0))=0. \tag{1.2}$$

这里 L 是 Lebesgue 测度(有的文献，把 L 换为其它固定的测度 λ，

此时可称 δ_0 是几乎(λ)容许估计).

这一章中讨论参数估计的容许性问题. 首先是讨论一维指数族及转移参数族(又称位置参数族)参数估计的容许性. 这是判决函数理论的重要问题之一. 自然, 我们希望所采用的估计是容许的. 这就引起研究以往常用估计的容许性. 可以想象过去认为优良的估计应当多数是容许的. 但有超乎人们想象的结果, 如 James 和 Stein 早在 1956 年及 1964 年就证明了三维以上的正态分布以样本均值估计均值, 在平方损失下是不容许的; 一维正态分布的方差, 在平方损失下样本方差的线性函数也是不容许的. 这就引起了人们的更大兴趣. 应当指出: 判定一个估计容许与否, 并非容易的事, 至今, 结果尚不多. 还要指出一点: 认为一个容许估计"就必然好"的观念是错误的. 例如, 若损失函数 $L(\theta, d)$ 对每个 $\theta\in\Theta$ 仅有唯一的 $d_\theta\in D$, 使得 $L(\theta, d)$ 在 d_θ 取极小, 则对某个 θ_0, 取 $\delta(x)=d_{\theta_0}$, 它显然是 θ 的容许估计, 但这个估计只顾到 θ_0 一点, 不顾到其余, 这是不会为人们所采用的. 下面再举一例:

例 1.1 令 $x_1, x_2, \cdots, x_n (n\geqslant 2)$ 是来自正态分布 $N(\mu, \sigma^2)$ 总体的简单抽样, 对于 $\theta=(\mu, \sigma^2)$, $\mu\in R'$, $\sigma>0$, 我们要估计 σ^2, $L(\theta, d)=(d\sigma^{-2}-1)^2$. 现在考查

$$\hat{\sigma}_1^2=(n-1)^{-1}\sum_{i=1}^n(x_i-\bar{x})^2, \quad R(\theta, \hat{\sigma}_1^2)=\frac{2}{n-1};$$

$$\hat{\sigma}_2^2=(n+1)^{-1}\sum_{i=1}^n(x_i-\bar{x})^2, \quad R(\theta, \hat{\sigma}_2^2)=\frac{2}{n+1};$$

$$\hat{\sigma}_3^2=\frac{1}{n+2}\sum_{i=1}^n x_i^2, \quad R(\theta, \hat{\sigma}_3^2)=\frac{2}{n+2}+\left(\frac{n}{n+2}\right)^2\frac{\mu^4}{\sigma^4}.$$

我们知道 $\hat{\sigma}_1^2$ 是 UMVUE, $\hat{\sigma}_2^2$ 是最优同变估计和 minimax 估计. 比较它们的风险, 有

$R(\theta, \hat{\sigma}_1^2)<R(\theta, \hat{\sigma}_2^2)$, 对一切 θ;

$R(\theta, \hat{\sigma}_3^2)\leqslant R(\theta, \hat{\sigma}_2^2)$, 对 $\lambda\leqslant 2(n+2)(n+1)^{-1}n^{-2}$;

$R(\theta, \hat{\sigma}_3^2)>R(\theta, \hat{\sigma}_2^2)$, 对 $\lambda>2(n+2)(n+1)^{-1}n^{-2}$.

这里 $\lambda=\mu^4/\sigma^4$. 注意 $R(\theta, \hat{\sigma}_3^2)\to\infty$ $(\lambda\to\infty)$, 看来 $\hat{\sigma}_3^2$ 是一个不好的估计, 但在后面要指出: $\hat{\sigma}_3^2$ 是容许估计, 而 $\hat{\sigma}_2^2$ 却不是容许

估计.

上面分析说明，单纯从容许性角度来选择估计是不尽合理的，因此，我们往往只考虑某类准则下优良估计的容许性问题. 最常见的是 minimax 估计或最优同变估计的容许性问题.

首先我们将给出证明容许性的一般定理.

定理 1.1 i) 若 δ_0 是关于损失函数 $L(\theta, d)$ 基本唯一(关于 $\mathscr{P}$)的 minimax 估计，则 δ_0 是容许的.

ii) 若 δ_0 是容许的，而且在 Θ 上为常数风险函数，则 δ_0 是 minimax 估计.

证 i) 若 δ_0 不是容许估计，则存在 $\delta_1 \prec \delta_0$，显然，$\delta_1(x) \neq \delta_2(x)$, a. s. P_θ，对某些 $\theta \in \Theta$ 成立. 然而 δ_1 也是 minimax 估计，与 minimax 估计的基本唯一性矛盾.

ii) 若 δ_0 不是 minimax 估计，则存在 δ_1 使得

$$R(\theta, \delta_1) \leqslant \sup_{\theta \in \Theta} R(\theta, \delta_1) < \sup_{\theta \in \Theta} R(\theta, \delta_0) = \text{const}$$

对 $\forall \theta \in \Theta$ 成立，这与 δ_0 的容许性矛盾.

定理 1.2(Blyth [2]) i) 设凡具有有限风险估计 δ，其风险函数 $R(\theta, \delta)$ 是 Θ 上的连续函数，$\xi(\theta)$ 是 Θ 上的先验分布，且在 Θ 的任意点的任一邻域 V 有 $\xi(V) > 0$，其对应的 Bayes 估计为 δ_ξ，其风险 $R(\xi, \delta_\xi) < \infty$，则 δ_ξ 是容许的.

ii) 如果 $\xi(\theta)$ 是 Θ 上的先验分布，其相应的 Bayes 估计 δ_ξ 是唯一的，则 δ_ξ 是容许估计.

证 ii) 若 δ_ξ 是非容许的，则存在 $\delta_1 \prec \delta_\xi$，可见 δ_1 也是 Bayes 估计，但与 δ_ξ 不同. 这与 Bayes 估计的唯一性假设矛盾.

i) 若 δ_ξ 是非容许的，则存在 $\delta_1 \prec \delta_\xi$，那么

$$R(\theta, \delta_1) \leqslant R(\theta, \delta_\xi), \quad 当 \quad \theta \in \Theta.$$

且在某点 $\theta_0 \in \Theta$，严格不等号成立. 根据连续性假设，存在 θ_0 的邻域 V_{θ_0}，在 $\theta \in V_{\theta_0}$ 上式严格不等号成立. 因为 $\xi(V_{\theta_0}) > 0$，故

$$\int_\Theta R(\theta, \delta_1) d\xi < \int_\Theta R(\theta, \delta_\xi) d\xi.$$

这与 δ_ξ 是 Bayes 估计矛盾.

注 结论对广义 Bayes 估计也成立，如广义 Bayes 风险有限.

上述定理虽然简单，但有实用价值，它产生了一个容许估计子类. 不少 minimax 估计或者属于此类，或者是这类估计的极限. 下面几节正好利用这点证明 minimax 估计的容许性. 但也要指出，并非所有这个 Bayes 估计类的极限是容许的，它涉及一个收敛速度问题. 此现象正好是研究容许估计的困难所在. 这里就不细说了，先举一例说明之.

续例 1.1 我们取 (μ, σ^2) 的先验分布为 $\mu \sim N(0, \tau^2)$，$\sigma^{-2} \sim G(\lambda, V)$（参数为 $\lambda>0$，$V>0$ 的 Γ-分布），且 μ 与 σ^2 独立. 记 $\bar{x}=\sum\limits_{i=1}^{n} x_i/n$，$s^2=\sum\limits_{i=1}^{n}(x_i-\bar{x})^2$，$\sigma^2$ 在损失函数 $(d\sigma^{-2}-1)^2$ 下的 Bayes 估计为

$$
\begin{aligned}
\hat{\sigma}^2_{\tau,\lambda,\nu}=&\left(\int_0^\infty y^{\nu+n/2}e^{-\lambda y}\,dy\int_{-\infty}^{\infty}\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\theta)^2 y\right.\right.\\
&\left.\left.-\theta^2\Big/2\tau^2\right\}d\theta\right)\Big/\left(\int_0^\infty y^{\nu+1+n/2}e^{-\lambda y}\,dy\right.\\
&\left.\times\int_{-\infty}^{\infty}\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\theta)^2 y-\theta^2\Big/2\tau^2\right\}d\theta\right)\\
=&\left(\int_0^\infty y^{\nu+n/2}e^{-\lambda y}\,dy\int_{-\infty}^{\infty}\exp\left\{-\frac{1}{2}(ny+\tau^{-2})\left(\theta-\frac{nyx}{ny+\tau^{-2}}\right)^2\right.\right.\\
&\left.\left.-\frac{1}{2}\sum_{i=1}^{n}x_i^2 y+\frac{1}{2}\frac{(ny\bar{x})^2}{ny+\tau^{-2}}\right\}d\theta\right)\Big/\int_0^\infty y^{\nu+1+n/2}e^{-\lambda y}\,dy\\
&\times\int_{-\infty}^{\infty}\exp\left\{-\frac{1}{2}(ny+\tau^{-2})\left(\theta-\frac{ny\bar{x}}{ny+\tau^{-2}}\right)^2\right.\\
&\left.-\frac{1}{2}\sum_{i=1}^{n}x_i^2 y+\frac{1}{2}\frac{(ny\bar{x})^2}{ny+\tau^{-2}}\right\}d\theta\\
=&\int_0^\infty y^{\nu+n/2}(1+ny\tau^2)^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\frac{n\bar{x}^2 y}{1+ny\tau^2}\right.\\
&\left.-y(s^2/4+\lambda)\right\}dy\Big/\int_0^\infty y^{\nu+1+n/2}(1+ny\tau^2)^{-\frac{1}{2}}\\
&\times\exp\left\{-\frac{1}{2}\frac{n\bar{x}^2 y}{1+ny\tau^2}-y(s^2/2+\lambda)\right\}dy. \qquad (1.3)
\end{aligned}
$$

根据定理 1.2 它是容许估计. 我们可以看到它不仅是 s^2 的函数, 也是 $\bar{x}^2$ 的函数. 众所周知, $\bar{x}$ 与 s^2 是独立的, 似乎 $\bar{x}$ 与 σ^2 无关, 事实上, 因为 $E\bar{x}^2=\mu^2+\sigma^2/n$, $\bar{x}$ 仍然提供了关于 σ^2 的信息. 这也就是常见到的仅仅依赖 s^2 的 σ^2 的估计并不容许的一个直观道理. 从上式可见到, 当 $\tau^2\to\infty$, $\lambda\to 0$, $\nu\to n/2$ 时 $\hat{\sigma}^2_{\tau,\lambda,\nu}\to s^2/(n+1)$. 在 §4 将证明 $s^2/(n+1)$ 不是 σ^2 的容许估计, 这就说明了 Bayes 估计的极限不一定是容许估计.

一个估计 $\delta_{\varepsilon,\xi}\in\mathscr{D}$ 称为关于 ξ、θ 的 ε-Bayes 估计, 如果

$$R(\xi,\ \delta_{\varepsilon,\xi})\leqslant\inf_{\delta\in\mathscr{D}}R(\xi,\ \delta)+\varepsilon. \tag{1.4}$$

C. Stein [26, i)] 利用这个概念, 建立了下列定理:

定理 1.3 若 $\delta_0\in\mathscr{D}$, 对任意的 $\theta_0\in\Theta$, $\varepsilon>0$, 存在 Θ 上一个先验分布 $\xi^{(1)}$ 及 $0<p<1$, 使得 δ_0 是关于 $\xi=(1-p)\xi^{(1)}+p\xi_{\theta_0}$、$\theta$ 的 εp-Bayes 估计, 则 δ_0 是 θ 的容许估计.

这里 ξ_{θ_0} 是概率集中 θ_0 点的单点分布.

证 若 δ_0 是非容许的, 则存在 $\delta_1\in\mathscr{D}$, 使得

$$R(\theta,\ \delta_1)\leqslant R(\theta,\ \delta_0),\quad \forall\theta\in\Theta.$$

且不等号对某些 $\theta\in\Theta$ 成立. 按照假设 δ_0 是关于 ξ、θ 的 εp-Bayes 估计, 故

$$R(\xi,\ \delta_0)\leqslant\inf_{\delta\in\mathscr{D}}R(\xi,\ \delta)+\varepsilon p\leqslant R(\xi,\ \delta_1)+\varepsilon p.$$

即

$$\begin{aligned}(1-p)R(\xi^{(1)},\ \delta_0)&+pR(\theta_0,\ \delta_0)\\ &\leqslant(1-p)R(\xi^{(1)},\ \delta_1)+pR(\theta_0,\ \delta_1)+\varepsilon p\\ &\leqslant(1-p)R(\xi^{(1)},\ \delta_0)+pR(\theta_0,\ \delta_1)+\varepsilon p\end{aligned} \tag{1.5}$$

$$\Rightarrow\qquad R(\theta_0,\ \delta_0)\leqslant R(\theta_0,\ \delta_1)+\varepsilon. \tag{1.6}$$

因为 $\theta_0\in\Theta$、ε 的任意性, 从而有 $R(\theta,\ \delta_0)\leqslant R(\theta,\ \delta_1)$ 对 $\forall\theta\in\Theta$ 成立, 这与 δ_1 的假设矛盾, 定理得证.

下面我们将继续引进几个概念:

δ_ξ 称为关于参数空间 Θ 上测度 ξ 的正规广义 Bayes 判决函数(记为 NGBDF), 如果

$$\int R(\theta,\ \delta_\xi)d\xi(\theta)=\inf_\delta\int R(\theta,\ \delta)d\xi(\theta).$$

显然,要求对所有 $\theta\in\Theta$、$R(\theta,\ \delta_\xi)<\infty$, 这才有意义.

δ_ξ 是关于参数空间 Θ 上测度 ξ 的广义 Bayes 判决函数(GBDF), 如果

$$E_\xi(L(\theta,\ \delta_\xi)\,|\,x)=\inf_{d\in D}E_\xi(L(\theta,\ d)\,|\,x) \tag{1.7}$$

或者

$$\int L(\theta,\ \delta_\xi)p(x,\ \theta)d\xi(\theta)=\inf_{d\in D}\int L(\theta,\ d)p(x,\ \theta)d\xi(\theta).$$

这里 $p(x,\ \theta)$ 是 $\frac{dP_\theta}{d\mu}$.

δ_ξ 称为关于参数空间 Θ 上测度 ξ 的比较广义 Bayes 判决函数(CGBDF), 若对任意判决函数 δ, 有

$$\Delta_\xi(\delta,\ \delta_\xi)=\int[R(\theta,\ \delta)-R(\theta,\ \delta_\xi)]d\xi(\theta)\geqslant 0. \tag{1.8}$$

则很容易证明 GBDF⇒NGBDF.

引理 1.1 设 ξ 为 $(\Theta,\ \mathscr{B}_\Theta)$ 上的一个测度. $\{\xi_i\}$ 是 $(\Theta,\ \mathscr{B}_\Theta)$ 上一串测度, 而且 $\liminf\limits_{i\to\infty}\xi_i(E)\geqslant\xi(E)$ 对一切 $E\in\mathscr{B}_\Theta$ 成立, $g(\theta)$ 是 $(\Theta,\ \mathscr{B}_\Theta)$ 上非负可测函数, 则

$$\liminf_{i\to\infty}\int g(\theta)d\xi_i(\theta)\geqslant\int g(\theta)d\xi(\theta).$$

证 任取一正数 M,

$$\begin{aligned}
&\liminf_{i\to\infty}\int g(\theta)d\xi_i(\theta)\\
&\quad\geqslant\liminf_{i\to\infty}\sum_{k=1}^{2^n}\frac{KM}{2^n}\xi_i\left(\theta:\frac{MK}{2^n}<g\leqslant\frac{M(K+1)}{2^n}\right)\\
&\quad\geqslant\sum_{k=1}^{2^n}\frac{KM}{2^n}\liminf_{i\to\infty}\xi_i\left(\theta:\frac{MK}{2^n}<g(\theta)\leqslant\frac{M(K+1)}{2^n}\right)\\
&\quad\geqslant\sum_{k=1}^{2^n}\frac{KM}{2^n}\xi\left(\theta:\frac{MK}{2^n}<g(\theta)\leqslant\frac{M(K+1)}{2^n}\right).
\end{aligned}$$

然后令 $n\to\infty$, 再令 $M\to\infty$, 由积分定义即得引理的结论.

定理 1.4 设 ξ 为 $(\Theta,\ \mathscr{B}_\Theta)$ 上的一个测度, δ_0 是 θ 的一个估

计，如果:

i) 当估计 $\delta_1 \prec \delta_0$ 时，则存在 $E \in \mathscr{B}_\Theta$，使 $\xi(E)>0$，$R(\theta, \delta_1) < R(\theta, \delta_0)$ （当 $\theta \in E$）.

ii) 在 $(\Theta, \mathscr{B}_\Theta)$ 上存在一串测度 $\{\xi_i\}$，使得

a) $\liminf\limits_{i\to\infty} \xi_i(T) \geqslant \xi(T)$ （当 $T \in \mathscr{B}_\Theta$）;

b) $\lim\limits_{i\to\infty} \Delta_{\xi_i}(\delta_0, \delta_{\xi_i}) = 0$.

此处 δ_{ξ_i} 是 ξ_i 的 CGBDF.

则 δ_0 是 θ 的容许估计.

证 若 δ_0 是非容许的，则存在 δ^*，使得 $\delta^* \prec \delta_0$. 由假设 i) 知 $R(\theta, \delta_0) - R(\theta, \delta^*) > 0$ （当 $\theta \in E$），且 $\xi(E)>0$,

$$\Delta_{\xi_i}(\delta_0, \delta^*) \geqslant \int_E [R(\theta, \delta_0) - R(\theta, \delta^*)] d\xi_i.$$

根据引理 1.1,

$$\liminf_{i\to\infty} \Delta_{\xi_i}(\delta_0, \delta^*) \geqslant \int_E [R(\theta, \delta_0) - R(\theta, \delta^*)] d\xi > 0,$$

$$\begin{aligned}\liminf_{i\to\infty} \Delta_{\xi_i}(\delta_0, \delta_{\xi_i}) &= \liminf_{i\to\infty} [\Delta_{\xi_i}(\delta_0, \delta^*) + \Delta_{\xi_i}(\delta^*, \delta_{\xi_i})] \\ &\geqslant \liminf_{i\to\infty} \Delta_{\xi_i}(\delta_0, \delta^*) > 0. \end{aligned} \tag{1.9}$$

这与假设 ii) 的 b) 矛盾，因此 δ_0 是容许的.

注 1) 如果 ξ 能使 i) 满足，δ_ξ 是 CGBDF，则 δ_ξ 是容许的. $\{\xi_i\}$ 往往取有限测度，此时 δ_{ξ_i} 往往是 Bayes 解.

2) 如果 Θ 上有拓扑结构，假设在 i)、ii) 中将 E 和 T 皆换为开集，定理结论仍然成立. 例如假设 $R(\theta, \delta_1)$、$R(\theta, \delta_0)$ 关于 θ 皆连续.

3) 如果 $L(\theta, d)$ 是严凸损失函数，而假设分布族满足：如有 $B \in \mathscr{B}_x$ 及 $\theta_0 \in \Theta$ 使得 $P_{\theta_0}(B)>0$，那么存在 $E \in \mathscr{B}_\Theta$ 使 $\xi(E)>0$，$P_\theta(B)>0$ （当 $\theta \in E$）. 此条件可替换假设 i) 使定理成立. 如果其中 E 还是 θ_0 的邻域，假设 ii) 中 T 可减弱为开集. 因为若有 $\delta_1 \prec \delta_0$，就有 $\theta_0 \in \Theta$，使对某个 $B \in \mathscr{B}_x$ 有 $P_{\theta_0}(B)>0$ 及 $\delta_1(x) \neq \delta_2(x)$ （当 $x \in B$）. 那么记 $\delta_2 = \frac{1}{2}(\delta_1 + \delta_0)$，有 $\delta_1 \prec \delta_2 \prec \delta_0$，而且由

假设知 $R(\theta,\ \delta_2)\leqslant R(\theta,\ \delta_1)$ (对一切 θ), 且 $\theta\in E$ 时取严格不等号. 因此上面的假设可代替 i).

4) 在某些假设下, 假设 ii) 可成为 δ_0 容许的必要条件. 读者可参阅 Farrel[12, ii)].

上述定理虽然简单, 但往往是求证一个估计容许性的重要方法.

下面来讨论一下几个独立参数混合而成为多参数时, 其原来的容许估计是否保持容许性的问题.

定理 1.5 假设 $X_j\sim P_{\theta_j}$, 记 $p_{\theta_j}(x_j)=\dfrac{dP_{\theta_j}}{d\mu_j}$, $\theta_j\in\Theta_j$ $(j=1,\ \cdots,\ k)$, $X_1,\ \cdots,\ X_k$ 相互独立. 对于 $j=1,\ \cdots,\ k$ 有如下假设:

i) 对 $(\mathscr{X}_j,\ \mathscr{B}_{x_j})$ 上某可测集 B_j 若有 $\theta_j^0\in\Theta_j$, 使 $P_{\theta_j^0}(B_j)>0$. 则一定存在 $E_j\in\mathscr{B}_{\Theta_j}$, 使得 $\xi_j(E_j)>0$, $P_{\theta_j}(B_j)>0$ (当 $\theta_j\in E_j$), 此处 ξ_j 为给定的 $(\Theta_j,\ \mathscr{B}_{\Theta_j})$ 上的 σ-有限测度.

ii) 在 $(\Theta_j,\ \mathscr{B}_{\Theta_j})$ 上有一串有限测度 $\{\xi_{ij}\}$ 及 σ-有限测度 ω_j, 使得 ξ_j、$\{\xi_{ij}\}$ 被 ω_j 所控制, 记 $g_j(\theta_j)=\dfrac{d\xi_j}{d\omega_j}$, $g_{ij}(\theta_j)=\dfrac{d\xi_{ij}}{d\omega_j}$ $(i=1,\ 2,\ \cdots)$, 并设 $\liminf\limits_{i\to\infty} g_{ij}(\theta_j)\geqslant g_j(\theta_j)$, a. s. ω_j.

iii) 估计 θ_j 的损失函数 $L(\theta_j,\ d_j)$ 是严凸的. 取 $V_j(\theta_j)>0$, $\theta_j\in\Theta_j$. 估计 $\theta=(\theta_1,\ \cdots,\ \theta_k)\in\Theta=\Theta_1\times\cdots\times\Theta_k$ 的损失函数取为 $\sum\limits_{j=1}^{k}V(\theta_j)L(\theta_j,\ d_j)$.

如果估计 $\delta_{j0}(x_j)$ $(j=1,\ \cdots,\ k)$ 满足:

$$\lim_{i\to\infty}\Delta_{\xi_{ij}}(\delta_{j0},\ \delta_{ij})=0\quad(j=1,\ \cdots,\ k);\tag{1.10}$$

$$\lim_{i\to\infty}\sum_{j=1}^{k}\Delta_{\xi_{ij}}(\delta_{j0},\ \delta_{\xi_{ij}})\int\prod_{\nu\neq j}^{k}(V_\nu^{-1}(\theta_\nu)d\xi_{i\nu}(\theta_\nu))=0.\tag{1.11}$$

那么 $\delta_0(x_1,\ \cdots,\ x_k)=(\delta_{10}(x_1),\ \cdots,\ \delta_{k0}(x_k))$ 是 θ 的容许估计. 这里 $\delta_{\xi_{ij}}$ 为 OGBDF.

证 记 $d\xi=d\xi_1\cdots d\xi_k$, $d\xi^{(i)}=d\xi_{i1}\cdots d\xi_{ik}$, $d\omega=d\omega_1\cdots d\omega_k$. 对任一 $B\in\mathscr{B}_{x_1\times x_2\times\cdots\times x_k}$, 则有

$$\liminf_{i\to\infty}\int_B \prod_1^k V_j(\theta_j)d\xi^{(i)}$$

$$=\liminf_{i\to\infty}\int_B \prod_{j=1}^k V_j(\theta_j)g_{ij}(\theta_j)d\omega$$

$$\geqslant\int_B \prod_1^k [V_j(\theta_j)\liminf_{i\to\infty} g_{ij}(\theta_j)]d\omega$$

$$\geqslant\int_B \left[\prod_1^k V_j(\theta_j)g_j(\theta_j)\right]d\omega. \tag{1.12}$$

因为ξ_j为$(\Theta_j, \mathscr{B}_{\Theta_j})$上的$\sigma$-有限测度，故存在彼此不相交的可测集$\{A_{jl}\}$，使得$\bigcup_{l=1}^{\infty} A_{jl}=\Theta_j$，$0<\xi_j(A_{jl})<\infty$. 现作概率测度$\bar{\xi}_j(F)=\sum_{l=1}^{\infty}\xi_j(F\cap A_{jl})(\xi_j(A_{jl}))^{-1}2^{-l}$（当$F\in\mathscr{B}_{\Theta_j}$）. 显然，$\xi_j$与$\bar{\xi}_j$ $(j=1, \cdots, k)$是等价的.

再记$\bar{P}_j(x_j)=\int p_{\theta_j}(x_j)d\bar{\xi}_j$，$\overline{\mathscr{X}}_j=\{x_j:\bar{P}_j(x_j)>0\}$，对任一$F\in\mathscr{B}_{\mathscr{X}_j}$. 若$F\cap\overline{\mathscr{X}}_j=\emptyset$，则$P_{\theta_j}(F)=0$（当$\theta_j\in\Theta_j$）. 否则，存在$\theta_j^0\in\Theta_j$，使$P_{\theta_j^0}(F)>0$. 根据假设存在$E_j\in\mathscr{B}_{\Theta_j}$，使$\xi_j(E_j)>0$，$P_{\theta_j}(F)>0$（对于$\theta_j\in E_j$）. 因而$\int_F\bar{P}_j(x_j)d\mu_j>0$，这与$F\cap\overline{\mathscr{X}}_j=\emptyset$ $(j=1, \cdots, k)$矛盾.

因此，若对某一$B\in\mathscr{B}_{\mathscr{X}_1\times\cdots\times\mathscr{X}_k}$存在$\theta^0\in\Theta$，使得

$$\int_B \prod_{j=1}^k p_{\theta_j^0}(x_j)d\mu_1\cdots d\mu_k>0,$$

那么
$$\int_{B\cap\overline{\mathscr{X}}_1\times\cdots\times\overline{\mathscr{X}}_k} \prod_{j=1}^k p_{\theta_j^0}(x_j)d\mu_1\cdots d\mu_k>0.$$

从而$\mu_1\times\cdots\times\mu_k(\widetilde{B})>0$，此处$\widetilde{B}=B\cap\overline{\mathscr{X}}_1\times\cdots\times\overline{\mathscr{X}}_k$. 故

$$\int_{\widetilde{B}} \prod_1^k \bar{P}(x_j)d\mu_1\cdots d\mu_k>0 \quad 及 \quad \int_B \prod_1^k \bar{P}(x_j)d\mu_1\cdots d\mu_k>0.$$

再根据Fubini定理知，存在$E\in\mathscr{B}_{\Theta_1\times\cdots\times\Theta_k}$，$\bar{\xi}(E)>0\Rightarrow\xi(E)>0\Rightarrow$

$$\int_E \prod_1^k V_{1j}(\theta_j)d\xi>0. \tag{1.13}$$

根据定理1.4的注3知，若$\delta_1\prec\delta_0$，则有E，使(1.13)式成立，而

且 $R(\theta, \delta_1) \leqslant R(\theta, \delta_0)$ (当 $\theta \in \Theta_1 \times \cdots \times \Theta_k$), 在 $\theta \in E$ 时, 严格不等号成立.

由 $X_1, X_2, \cdots, X_k$ 相互独立, 又假定 $\xi_{i1}, \cdots, \xi_{ik}$ 为有限测度, $\xi^{(i)} = (\xi_{i1}, \cdots, \xi_{ik})$, 故 $\delta_{\xi_{ij}}(x_j)$ 是 θ_j 在 $L(\theta_j, d_j)$ 下的 CGBDF, 那么 $\delta_{\xi^{(i)}} = (\delta_{\xi_{i1}}, \cdots, \delta_{\xi_{ik}})$ 也是 θ 在 $\sum\limits_{j=1}^{n} V(\theta_j) L(\theta_j, d_j)$ 下关于 $\prod\limits_{1}^{k} V(\theta_j) d\xi_{i1} \cdots d\xi_{ik}$ 下的 CGBDF. 故(1.11)相当于定理 1.4 的假设 ii) 的 b).

上面的论述已经验证定理 1.4 的条件成立, 因此 δ_0 是 θ 的容许估计.

由定理 1.4 也知 $\delta_{j0}(x_j)$ 是在 $L(\theta_j, d_j)$ 及 $(x_j, \mathscr{B}_j)$ 限制下 θ_j 的容许估计. 但 $\delta_{j0}(x_j)$ 容许与否, 和 $V_j(\theta_j) > 0$ 的选择无关. 定理是说明, 在一定条件下, $\delta_j(x_j)$ 是 θ_j 的容许估计(损失函数为 $L(\theta_j, d_j)$), 那么 $\delta^0(x) = \{\delta_1(x_1), \cdots, \delta_k(x_k)\}'$ 是在 $\prod\limits_{j=1}^{k} L(\theta_j, d_j)$ 下的容许估计. 此定理的应用, 请参见[9, ii)].

例 1.2[7, iii)] 若 $X_j \sim N(\theta_j, 1)$, $\theta_j \in R$ $(j=1, \cdots, k)$, 且 $X_1, \cdots, X_k$ 相互独立, 则 $X = (X_1, \cdots, X_k)'$ 在损失函数

$$\sum_{j=1}^{k} (1+\theta_j^2)^{r/2} (\theta_j - d_j)^2$$

下, 当 $r \geqslant 1$ 时以及 $r < 1$、$k < (2-r)/(1-r)$ 时, x 是 $\theta = (\theta_1, \cdots, \theta_k)'$ 的容许估计.

取 $g_{ij}(\theta_j) = e^{-\theta^2/i}$ $(i=1, 2, \cdots)$, $g_j(\theta_j) = 1$; $d\omega_j = d\theta_j$. 它们显然满足定理 1.5 的假设 i)~iii), 现在主要验证(1.10)与(1.11)式是否成立.

在平方损失下, θ_j 关于 ξ_{ij} 的 Bayes 估计 $\delta_{ij} = \dfrac{i}{1+i} x_j$,

故 $$\Delta_{\xi_{ij}}(x_j, \delta_{\xi_{ij}}(x_j)) = \frac{\sqrt{2\pi i}}{1+i} \to 0 \quad (i \to \infty).$$

而 $$\int_{-\infty}^{\infty} (1+\theta^2)^{-\frac{r}{2}} e^{-\frac{t^2}{2i}} d\theta \to \int_{-\infty}^{\infty} (1+\theta^2)^{-\frac{r}{2}} d\theta \quad (\text{当 } r > 1),$$

$$\int_{-\infty}^{\infty}(1+\theta^2)^{-\frac{r}{2}}e^{-\frac{\theta^2}{2i}}d\theta\leqslant\int_{-\infty}^{\infty}|\theta|^{-r}e^{-\frac{\theta^2}{2i}}d\theta$$

$$=i^{\frac{1-r}{2}}\int_{-\infty}^{\infty}|\theta|^{-r}e^{-\theta^2/2}d\theta\quad (\text{当 } r<1)$$

$$\int_{-\infty}^{\infty}(1+\theta^2)^{-\frac{1}{2}}e^{-\theta^2/2i}d\theta$$

$$\leqslant\int_{|\theta|\leqslant 1}(1+\theta^2)^{-\frac{1}{2}}e^{-\theta^2/2i}+\int_{|\theta|>i^{-1/2}}\theta^{-1}e^{-\theta^2/2}d\theta$$

$$=O(\log i).$$

因此,

$$\lim_{i\to\infty}\sum_{j=1}^{k}\varDelta_{\xi_{ij}}(x_j,\ \delta_{\xi_{i,j}}(x_j))\int_{R^{k-1}}\prod_{\nu\neq j}^{k}(1+\theta_\nu^2)^{-r/2}e^{-\theta_\nu^2/2i}d\theta$$

$$=\lim_{i\to\infty}k\frac{\sqrt{2\pi i}}{1+i}O(i^{\frac{(k-1)(1-r)^+}{2}}\log i)=0.$$

这里$(1-r)^+=1-r$(当 $r<1$), $(1-r)^+=0$(当 $r\geqslant 1$).

由定理 1.5 知例 1.2 的结论成立.

Brown[7, 111] 还进一步证明了在 $r<1$ 时, $k>(2-r)/(1-r)$ 时, X 是 θ 的非容许估计. 这说明容许性与损失函数的选取密切相关. 稍微变化一下, 损失函数就可能使原来的容许估计变成非容许的了.

例 1.3 令 X_i 具有 0、1 两点分布, $P(X_i=1)=p_i$, $0<p_i<1$. 且 $X_1, X_2, \cdots$ 相互独立. 那么对任意 $V_i(p_i)>0(0<p_i<1)$, $X^{(k)}=\frac{1}{4}(2X_1+1, \cdots, 2X_k+1)'$ 皆为在损失函数 $\sum_{i=1}^{k}V_i(p_i)\times(p_i-d_i)^2$ 下 $p^{(k)}=(p_1, \cdots, p_k)$的容许估计.

容易算得,若取 $\delta_{i0}(0)=\frac{1}{4}$, $\delta_{i0}(1)=\frac{3}{4}$, 它是$\xi_i\left(p_i=\frac{1}{2}-\frac{1}{\sqrt{2}}\right)=\xi_i\left(p_i=\frac{1}{2}+\frac{1}{\sqrt{2}}\right)=\frac{1}{2}$在平方损失$(d_i(x_i)-p_i)^2$下的 Bayes 估计, 因为 $E(\delta_{i0}(x_i)-p_i)^2=\frac{1}{16}$, 故它是 minimax 估计.

若取 $d\xi=\prod_1^k V_i^{-1}(p_i)d\xi_1\cdots d\xi_k\Big/\int\prod_1^k V_i^{-1}(p_i)d\xi_1\cdots d\xi_k.$

那么$(\delta_1(x_1), \cdots, \delta_k(x_k))'$ 也是关于 ξ 在损失函数

$$\sum_{i=1}^{k} V_i(p_i)(p_i-d_i)^2$$

下的 Bayes 估计．因此也是容许的，如果损失函数换为

$$\sum_{i=1}^{k} V_i(p_i)(p_i-d_i)^2 \Big/ \sum_{1}^{k} V_i(p_i),$$

那么$(\delta_1(x_1),\ \cdots,\ \delta_k(x_k))'$ 还是 $p^{(k)}$ 的容许 minimax 估计．

此例说明，无论怎样选择 $V_i(p_i)$，$(\delta_1(x_1),\ \cdots,\ \delta_k(x_k))'$ 的容许性不变．它完全不同于正态分布情况．

在第一章 §1 中，我们已介绍了判决函数的基本完全类和完全类的概念．现在我们再引入极小完全类的概念．若 $C^* \subset \mathscr{D}$ 是两个估计类，并且 C^* 为 $\mathscr{D}$ 中的完全类．但 C^* 的任一真子集都不是 $\mathscr{D}$ 中的完全类，则称 C^* 为关于 $\mathscr{D}$ 的极小完全类．当然，这些概念都是在损失函数给定下而言的．

令 $\mathscr{A}$ 表示(在损失函数给定下)所有的容许估计组成的类．如果极小完全类存在，它就与 $\mathscr{A}$ 重合．这是因为 $\mathscr{A}$ 一定包含在所有完全类中，本身不存在子完全类，而且如果极小完全类有一估计是非容许的，则去掉此估计，也组成一个完全类，这与极小完全类矛盾．从而就证明了我们叙述的结论．但应注意，极小完全类并不一定存在．下面举一例说明之．

例 1.4 随机变量 X 仅取 1、2 两点：

$$P(X=1)=\theta;\quad P(X=2)=1-\theta\ \ (0<\theta<1).$$

行动空间 $D=\{1,\ 2,\ \cdots\}$，$L(\theta,\ 1)=1-\theta$，$L(\theta,\ n)=n^{-1}$ (当 $n>1$)．令判决函数 δ_{ij} 为：$\delta_{ij}(1)=i$，$\delta_{ij}(2)=j$，当 $i>1$ 或 $j>1$ 时，δ_{ij} 显然是非容许的，只有 $\delta(1)=\delta(2)=1$ 时，才是容许的．$R(\theta,\ \delta_{11})=1-\theta$，但 δ_{11} 组不成完全类，因为对于 i、$j>1$，

$$1-\theta\leqslant\frac{\theta}{i}+\frac{1-\theta}{j}=R(\theta,\ \delta_{ij})$$

不能对所有的 $\theta\in(0,\ 1)$成立．

Wald 在他的著名著作《统计判决函数》中，给出了在一定条件下，极小完全类的存在性定理及所有广义 Bayes 估计(包括 Bayes 估计)组成一个基本完全类．

§2 平方损失下估计的容许性

一、概　　述

本节皆假设 Θ 是实轴上的一个区间 (a, b)，求可测实函数 $g(\theta)$ $(\theta\in\Theta)$ 的点估计，其损失函数采用平方损失 $\lambda(\theta)(g(\theta)-d)^2$，$\lambda(\theta)>0$, $\forall\theta\in\Theta$. 在求证 $g(\theta)$ 的某个估计的容许性时，与 $\lambda(\theta)$ 取何值无关，但对于讨论 minimax 估计，则有很大关系. 由于我们最感兴趣的是 minimax 估计的容许性，自然涉及 $\lambda(\theta)$ 的形式.

由于平方损失的特点，往往把估计的容许性问题转化为解常微分方程不等式的问题. 这一方法首先是50年代 Hodges, Lehmann[16], Girshick 和 Savage[14] 等人在处理某些指数型分布族均值的容许 minimax 估计时，用 Cramér-Rao 不等式把问题转化为解微分方程不等式，然后，Karlin 综合为普遍性定理. 同时用这个方法处理了 $p(x, \theta)=g(\theta)r(x)$ 截断参数族参数估计的容许性. 1961 年 Katz 又把这个方法作了综合，并对在限制参数域下对均值估计的容许性作了讨论. 近来又有新的发展. 例如 Zidek[28] 和 Brown[7,11] 的工作. 在本节中我们就用这个方法来讨论上面提到的问题. 请注意：由于平方损失是严凸损失函数，不失去一般性，本节皆设 $\delta(x)$ 是非随机化估计.

引理 2.1　假设 $Q(\theta)>0$, $T(\theta)\geqslant 0$ 为 (a, b) 上的可测函数，且当 $a<c<b$ 时，皆有

$$\int_c^x Q^{-1}(\theta)d\theta\to\infty \quad (\text{当 } x\to b); \tag{2.1}$$

$$\int_x^c Q^{-1}(\theta)d\theta\to\infty \quad (\text{当 } x\to a). \tag{2.2}$$

如果 $T(\theta)$ 满足不等式

$$\int_{t_1}^{t_2} T(\theta)Q(\theta)d\theta\leqslant k[Q(t_1)T^{1/2}(t_1)+Q(t_2)T^{1/2}(t_2)] \tag{2.3}$$

对所有 $a<t_1<t_2<b$ 成立（式中 k 为常数）.

则 $T(\theta)=0$, a. s. $[L]$ （L 为 Lebesgue 测度）.

证　i) 如果 $\liminf_{t_2\to b} Q(t_2)T^{1/2}(t_2)=\Delta>0$, 则对任意固定的 $c\in(a, b)$, 根据(2.3), 存在常数 k', $c_1>c$, 使得

$$H(t_2)\triangleq\int_c^{t_2} T(\theta)Q(\theta)d\theta\leqslant K'Q(t_2)T^{1/2}(t_2)\ (\text{当 } c_1\leqslant t_2<b).$$

由于 $H'(t)=T(t)Q(t)$, a. s. L. 因而

$$H^2(t_2)\leqslant K'^2Q^2(t_2)T(t_2)=K'^2Q(t_2)H'(t_2)\ (\text{当 } c_1\leqslant t_2<b).$$

若有 $c_0>c$, 使 $H(c_0)>0$, 由于 $H(t_2)$是非降的,不妨假定 $c_0>c_1$, 则由上式得

$$Q^{-1}(t_2)\leqslant K'^2H'(t_2)/H^2(t_2)=K'^2[-H^{-1}(t_2)]'$$
$$(\text{当 } c_0\leqslant t_2<b).$$

由这个微分不等式即得

$$\int_t^{t_2} Q^{-1}(\theta)d\theta\leqslant K'^2[H^{-1}(t)-H^{-1}(t_2)]$$
$$(\text{当 } c_0<t<b_2<b). \tag{2.4}$$

但根据假设 $t_2\to b$ 时, (2.4)式左端趋于无穷, 而右端趋向有限值 ($H(t_2)$非降),这就产生了矛盾,故

$$H(t_2)\equiv0\quad(\text{当 } c_1\leqslant t_2<b),$$

从而

$$T(\theta)=0,\quad \text{a. s.}\quad (c<\theta<b).$$

又由于 c 的任意性,故 $T(\theta)=0$ a. s. L $(a<\theta<b)$.

ii) 如果 $\liminf_{t_2\to b} Q(t_2)T^{1/2}(t_2)=0$, 则由 (2.3) 式可知对于 $a<t<b$, 有

$$G(t)\triangleq\int_t^b Q(\theta)T(\theta)d\theta\leqslant kQ(t)T^{\frac{1}{2}}(t)=k[-Q(t)G'(t)]^{1/2}.$$

如果存在一点 $t_0\in(a, b)$使得 $G(t_0)>0$, 则由 $G(t)$的非增性, 解微分不等式,得

$$k^2[G^{-1}(t_0)-G^{-1}(t)]\geqslant\int_t^{t_0} Q^{-1}(\theta)d\theta\quad(a<t<t_0). \tag{2.5}$$

根据(2.2)式, 当 $t\to a$ 时, 上式右端趋于正无穷大, 而左端为有限值, 这是矛盾的, 故 $G(t)\equiv0$ $(a<t<b)$. 从而 $T(\theta)=0$, a. s. L (当 $a<\theta<b$), 至此引理得证.

定理2.1 令 $Q(\theta)$ 满足 (2.1) 与 (2.2)，如果 $g(\theta)$ 的估计 $\delta(x)$ 的风险 $R(\theta, \delta)$ 在 (a, b) 上有限，而且存在可测函数 $Q_1(\theta)$，使之对所有的 $x\in\mathscr{X}$ $(a<t_1<t_2<b)$，有

$$\left|\int_{t_1}^{t_2}(\delta(x)-g(\theta))Q(\theta)Q_1(\theta)p(x, \theta)d\theta\right|$$

$$\leqslant k[Q(t_2)Q_1^{1/2}(t_2)p(x, t_2)+Q(t_1)Q_1^{1/2}(t_1)p(x, t_1)], \quad (2.6)$$

这里 $p(x, \theta)=dP_\theta/d\mu$ (μ 为 σ 有限测度)，则 $\delta(x)$ 是 $g(\theta)$ 的几乎容许估计.

证 若有 δ^*，使得当 $a<\theta<b$ 时，皆有

$$R(\theta, \delta^*)=\lambda(\theta)\int_x(\delta^*(x)-g(\theta))^2p(x, \theta)d\mu$$

$$\leqslant R(\theta, \delta)=\lambda(\theta)\int_x(\delta(x)-g(\theta))^2p(x, \theta)d\mu, \quad (2.7)$$

则利用熟知的平方展开式

$$(\delta^*(x)-g(\theta))^2$$

$$=(\delta^*(x)-\delta(x))^2+2(\delta^*-\delta)(\delta-g(\theta))+(\delta-g(\theta))^2$$

及 (2.7) 式推得对任意 $\theta\in(a, b)$，有

$$T(\theta)\triangleq\int_x(\delta^*(x)-\delta(x))^2p(x, \theta)d\mu$$

$$\leqslant 2\int_x(\delta(x)-\delta^*(x))(\delta(x)-g(\theta))p(x, \theta)d\mu.$$

从而由上式和 (2.6) 式，得

$$\int_{t_1}^{t_2}T(\theta)Q(\theta)Q_1(\theta)d\theta\leqslant 2\int(\delta(x)-\delta^*(x))$$

$$\times\left[\int_{t_1}^{t_2}(\delta(x)-g(\theta))Q(\theta)Q_1(\theta)p(x, \theta)d\theta\right]d\mu$$

$$\leqslant 2K\int|\delta(x)-\delta^*(x)|$$

$$\times[Q(t_2)Q_1^{1/2}(t_2)p(x, t_2)+Q(t_1)Q_1^{1/2}(t_1)p(x, t_1)]d\mu$$

$$\leqslant 2K[Q(t_2)Q_1^{1/2}(t_2)T^{1/2}(t_2)+Q(t_1)Q_1^{1/2}(t_1)T^{1/2}(t_1)]$$

(当 $a<t_1<t_2<b$).

上式最后一步是利用 Schwarz 不等式. 对上述不等式应用引理 2.1 知, $Q_1(\theta)T(\theta)=0$, a. s. L. 由此 $T(\theta)=0$, a. s. L 在 $T(\theta)=0$ 时可得 $R(\theta, \delta)=R(\theta, \delta^*)$. 这说明了 $\delta(x)$ 是几乎容许估计.

注 假设 $P_\theta(x)$ 满足对任意的 $\theta_0\in\Theta$, 若有 $B\in\mathscr{B}_x$ 满足 $P_{\theta_0}(B)>0$, 则存在 Θ 的子集 A, 使之 $L(A)>0$ 及 $P_\theta(B)>0$ (当 $\theta\in A$), 则几乎容许估计 δ 是容许估计.

证 若 δ 是非容许的, 则存在 $\delta^*\prec\delta$, 至少存在某个 $\theta_0\in\Theta$, $R(\theta_0, \delta^*)<R(\theta_0, \delta)$. 令 $B=\{x:\delta\neq\delta^*\}$, 作 $\delta^{**}=(\delta+\delta^*)/2$, 注意平方损失是严凸损失:

$$(\delta^{**}-g)^2\leqslant[(\delta-g)^2+(\delta^*-g)^2]/2,$$

且在 $x\in B$ 时, 严格不等号成立. 因为 $P_{\theta_0}(B)>0$, 由假设知存在 $A\subseteq\Theta$, $P_\theta(B)>0$ (当 $\theta\in A$ 时); $R(\theta, \delta^{**})<R(\theta, \delta)$ (当 $\theta\in A$ 时). 这与 δ 为几乎容许相矛盾. 故 δ 为容许估计.

二、指数族均值的容许估计

定理 2.2(Karlin) 若随机变量 X 服从指数族分布, 关于 σ 有限测度 μ 的密度函数为

$$p(x, \theta)=\beta(\theta)\exp[\theta T(x)] \quad (a<\theta<b). \tag{2.8}$$

若对任意 $c\in(a, b)$ 存在常数 $\lambda>0$ 及 k, 使之

$$\int_c^{t_2}\beta^{-\lambda}(\theta)e^{-k\lambda\theta}\,d\theta\to\infty \quad (\text{当 } t_2\to b);$$

$$\int_{t_1}^{c}\beta^{-\lambda}(\theta)e^{-k\lambda\theta}\,d\theta\to\infty \quad (\text{当 } t_1\to a).$$

则 $[T(x)+k\lambda]/(1+\lambda)$ 是 $E_\theta T(x)=g(\theta)$ 的容许估计.

证 取 $Q(\theta)=\beta^\lambda(\theta)e^{\lambda k\theta}$, $Q_1(\theta)=1$, 由于

$$g(\theta)=-\beta'(\theta)/\beta(\theta),$$

故

$$\left|\int_{t_1}^{t_2}\left[\frac{T(x)+k\lambda}{1+\lambda}+\frac{\beta'(\theta)}{\beta(\theta)}\right]Q(\theta)\beta(\theta)e^{\theta T(x)}\,d\theta\right|$$

$$= \left| \frac{1}{1+\lambda}\int_{t_1}^{t_2} \beta^{1+\lambda}(\theta) de^{\theta(T(x)+k\lambda)} + \int_{t_1}^{t_2} \beta^{\lambda}(\theta) e^{\lambda k\theta + T(x)\theta}\beta'(\theta) d\theta \right|$$
$$= \left| (1+\lambda)^{-1}\beta^{1+\lambda}(\theta) e^{\theta(T(x)+k\lambda)} \Big|_{t_1}^{t_2} \right|$$
$$\leqslant (1+\lambda)^{-1}[Q(t_2)p(x, t_2)+Q(t_1)p(x, t_1)].$$

因此定理 2.1 的条件成立. 故$[T(x)+k\lambda]/(1+\lambda)$是$g(\theta)$的几乎容许估计. 再由于$p(x, \theta)$有共同支撑, 由定理 2.1 的注知, 定理的结论成立.

我们利用定理 1.1 及定理 2.2 来证明下列的有关指数族均值某估计量的 minimax 容许性.

例 2.1 若损失函数取$(g(\theta)-d)^2/\sigma^2(\theta)$(此处$\sigma^2(\theta)$指指数族分布的方差), 同时有随机独立抽样的结果$x_1, x_2, \cdots, x_n$, 则有下述结果:

1) 样本均值 $\bar{x}=\frac{1}{n}\sum_{i=1}^{n} x_i$ 是正态分布$N(\mu, 1)$的均值μ的容许 minimax 估计. 根据上章 § 3 的结果, 当限制$\mu\geqslant\mu_0$时, $\bar{x}$ 也是 minimax 估计.

2) $\bar{x}$ 是 Poisson 分布均值 λ 的容许 minimax 估计.

3) $\bar{x}$ 是二项分布均值 p 的容许 minimax 估计.

4) 若损失函数为$(d-p)^2$, 则$(x+\sqrt{n}/2)/(n+\sqrt{n})$是二项分布均值 p 的容许 minimax 估计.

5) $\delta=\sum_{i=1}^{n}x_i^2\Big/(n+2)$是正态分布$N(0, \sigma^2)$的方差$\sigma^2$的容许 minimax 估计. 还应指出, 对于$N(\mu, \sigma^2)$, 在$\mu$及$\sigma^2$皆未知时, 上述$\delta$仍为$\sigma^2$的容许估计. 这是因为, 若有另外的估计$\delta^*$, 使

$$E\{(\delta^*-\sigma^2)^2|\mu, \sigma^2\}\leqslant E\{(\delta-\sigma^2)^2|\mu, \sigma^2\}$$

对 $-\infty<\mu<\infty$, $\sigma>0$ 成立, 必然也对 $\mu=0$ 成立. 从定理 2.1 的注即知: $\delta^*=\delta$ a. s. L, 从而证明 δ 是容许估计. 但在例 1.1 中已知 δ 的风险为 $\frac{2}{n+2}+\left(\frac{n}{n+2}\right)^2\left(\frac{\mu}{\sigma}\right)^4$ 当$|\mu/\sigma|\to\infty$时趋于∞. 因此, δ 不是 minimax 估计. 从上述推导中, 说明在损失函数为$(\sigma^2+\mu^2-d)^2/[\sigma^2(\sigma^2+\mu^2)+\mu^4(n+2)]$时, 上述 δ 是二阶矩

$\sigma^2+\mu^2$ 的容许 minimax 估计.

三、截断参数分布族参数的容许 minimax 估计

现在考虑一类截断参数分布族参数的估计，此分布的密度函数为

$$p(x,\ \theta)=\begin{cases}q(\theta)r(x), & 0<x<\theta,\ r(x)>0;\\ 0, & \text{其它}.\end{cases} \tag{2.9}$$

此处 $\Theta=(0,\ \infty)$, $q^{-1}(\theta)=\int_0^\theta r(x)dx<\infty$, $\int_0^\infty r(x)dx=\infty$. 我们的目的是估计 $q^{-\alpha}(\theta)$ $(0<\alpha<\infty)$. 易求出它的 UMVUE 为 $(\alpha+1)q^{-\alpha}(x)$, 由于

$$\begin{aligned}&E_\theta(cq^{-\alpha}(x)-q^{-\alpha}(\theta))^2\\&=(c^2(2\alpha+1)^{-1}-2c(\alpha+1)^{-1}+1)q^{-2\alpha}(\theta).\end{aligned}$$

显然，当 $c=(2\alpha+1)(\alpha+1)^{-1}$ 时，使上式达到最小. 这说明 UMVUE $(\alpha+1)q^{-\alpha}(x)$ 不是 $q^{-\alpha}(\theta)$ 的容许估计. 而 $(2\alpha+1)(\alpha+1)^{-1}q^{-\alpha}(x)$ 这估计量的容许与否，由下面定理回答.

定理 2.3 对分布族(2.9)在平方损失下，

$$\delta_0(x)=(2\alpha+1)(\alpha+1)^{-1}q^{-\alpha}(x)$$

是 $q^{-\alpha}(\theta)$ 的容许估计. 若损失函数取为 $q^{2\alpha}(\theta)(q^{-\alpha}(\theta)-d)^2$, 则 δ_0 还是 $q^{-\alpha}(\theta)$ 的容许 minimax 估计.

证 利用定理 2.1 来证: 令 $Q(\theta)=q(\theta)|q'(\theta)|^{-1}$, $Q_1(\theta)=(q'(\theta))^2q^{2\alpha-2}(\theta)>0$, $\beta\triangleq(2\alpha+1)(\alpha+1)^{-1}$, 由 $q(\theta)$ 的定义知

$$\begin{aligned}\int_{t_1}^{t_2}Q^{-1}(\theta)d\theta&=\int_{t_1}^{t_2}r(\theta)\left(\int_0^\theta r(x)dx\right)^{-1}d\theta\\&=\ln\int_0^\theta r(x)dx\Big|_{t_1}^{t_2}\to\infty\end{aligned}$$

(当 $t_2\to\infty$ 或 $t_1\to 0$ 时).

当 $0<x<t_1$ 时，利用 $q^{-\alpha}(x)$ 的非降性，得

$$\begin{aligned}&\left|\int_{t_1}^{t_2}[\beta q^{-\alpha}(x)-q^{-\alpha}(\theta)]p(x,\ \theta)Q(\theta)Q_1(\theta)d\theta\right|\\&\leqslant(1+\beta)\int_{t_1}^{t_2}r(x)q^{\alpha}(\theta)|q'(\theta)|d\theta\end{aligned}$$

$$\leqslant (1+\beta)(1+\alpha)^{-1} r(x)[q^{1+\alpha}(t_1)+q^{1+\alpha}(t_2)]$$
$$\leqslant (1+\beta)(1+\alpha)^{-1}[Q(t_1)Q_1^{1/2}(t_1)p(x,\ t_1)$$
$$+Q(t_2)Q_1^{1/2}(t_2)p(x,\ t_2)].$$

当 $t_1<x\leqslant t_2$ 时，注意，$p(x,\ \theta)$ 在 $0<x\leqslant\theta$ 才有正密度，故

$$\left|\int_{t_1}^{t_2}[\beta q^{-\alpha}(x)-q^{-\alpha}(\theta)]p(x,\ \theta)Q(\theta)Q_1(\theta)d\theta\right|$$
$$=\left|\int_{x}^{t_2}[\beta q^{-\alpha}(x)-q^{-\alpha}(\theta)]p(x,\ \theta)Q(\theta)Q_1(\theta)d\theta\right|$$
$$=\left|-(\alpha+1)^{-1}q^{-\alpha}(x)q^{2\alpha+1}(\theta)\Big|_x^{t_2}\right.$$
$$\left.+r(x)(\alpha+1)^{-1}q^{1+\alpha}(\theta)\Big|_x^{t_2}\right|$$
$$=\left|(\alpha+1)^{-1}q^{1+\alpha}(t_2)[1-q^{-\alpha}(x)q^{\alpha}(t_2)]\right|r(x)$$
$$\leqslant(\alpha+1)^{-1}q^{1+\alpha}(t_2)r(x)$$
$$=(\alpha+1)^{-1}Q(t_2)Q_1^{1/2}(t_2)p(x,\ t_2)$$
$$\leqslant(1+\beta)(1+\alpha)^{-1}[Q(t_1)Q_1^{1/2}(t_1)p(x,\ t_1)$$
$$+Q(t_2)Q_1^{1/2}(t_2)p(x,\ t_2)].$$

因此，定理 2.1 及其注的条件成立，故 $\delta_0=\beta q^{-\alpha}(x)$ 是 $q^{-\alpha}(\theta)$ 的容许估计，其平方损失的风险为 $\alpha^2(\alpha+1)^{-2}q^{-2\alpha}(\theta)$. 若损失函数取 $q^{2\alpha}(\theta)(q^{-\alpha}(\theta)-d)^2$, $\beta q^{-\alpha}(x)$ 的风险函数为常数，由定理 1.1 知它是 minimax 估计，定理得证.

例 2.2

$$p(x,\ \theta)=\begin{cases} nx^{n-1}/\theta^n, & 0\leqslant x<\theta \\ 0, & 其它 \end{cases}\quad (0<\theta<\infty).$$

它是 $(0,\ \theta)$ 上均匀分布的 n 次抽样结果顺序统计量中最大值的分布密度，现在估计 θ, 则在损失函数 $(\theta-d)^2\theta^{-2}$ 下，$\hat{\theta}=(n+2)(n+1)^{-1}x$ 是容许 minimax 估计.

用同样办法，可证与定理 2.3 类似的定理.

定理 2.4 若分布密度为

$$p(x,\ \theta)=\begin{cases} q(\theta)r(x), & 当\quad x\geqslant\theta,\ r(x)>0; \\ 0, & \theta_0\leqslant x<\theta. \end{cases}$$

此处 $\theta_0\leqslant\theta<\infty$, $q^{-1}(\theta)=\int_{\theta}^{\infty}r(x)dx$, $\lim\limits_{\theta\to\theta_0}q(\theta)=0$.

则在损失函数 $q^{2\alpha}(\theta)(q^{-\alpha}(\theta)-d)^2$ 下，$(2\alpha+1)(\alpha+1)^{-1}q^{-\alpha}(x)$ 是 $q^{-\alpha}(\theta)$ 的容许 minimax 估计.

证明在此略，留给读者作练习.

例 2.3

$$p(x,\ \theta)=\begin{cases} n\exp[n(\theta-x)], & 当\ \theta\leqslant x<\infty; \\ 0, & 其它. \end{cases} \quad (-\infty<\theta<\infty)$$

则在 $e^{2\theta}(e^{-\theta}-d)^2$ 损失下 $(n+2)(n+1)^{-1}e^{-x}$ 是 $e^{-\theta}$ 的容许 minimax 估计，此例在可靠性中有重要应用.

四、限制参数空间的容许估计问题

下面讨论限制参数空间的容许估计问题. 例如指数族分布 $dP_\theta/d\mu=p(x,\ \theta)=\beta(\theta)\exp[\theta T(x)]$，自然参数空间为 $(a,\ b)=\{\theta:0<\beta(\theta)<\infty\}$，$a$ 可取 $-\infty$；b 可取 $+\infty$，a、b 可以包含在区间内，也可以不包含在区间内. 有时，我们只需要考虑部分参数空间的情况. Katz 建立了一个一般性的定理，讨论了在 $\Theta=\{\theta:\theta\geqslant\theta_0,\ \theta\in(a,\ b)\}$ $(a<\theta_0<b)$ 指数族均值 $g(\theta)=-\beta'(\theta)\beta^{-1}(\theta)$ 的容许估计问题.

引理 2.2 设 $Q(\theta)>0$, $T(\theta)\geqslant0$, 为 $(\theta_0,\ b)$ 或 $(a,\ \theta_0)$ 上的可测函数，而且

$$\int_{\theta_0}^{x}Q^{-1}(\theta)d\theta\to\infty(-\infty)\quad 当\ x\to b(a).$$

若对某个常数 k, $T(\theta)$ 满足

$$\left|\int_{\theta_0}^{t}T(\theta)Q(\theta)d\theta\right|\leqslant kQ(t)T^{1/2}(t)\quad(\theta_0\leqslant t<b\ 或\ a<t\leqslant\theta_0).$$

则 $T(\theta)=0$, a. s. L, $(\theta\in(\theta_0,\ b)$ 或 $\theta\in(a,\ \theta_0))$.

证明与引理 2.1 类似，在此略.

定理 2.5 若 $\mathscr{P}=\{P_\theta:\ \theta\in(a,\ b)\}$, $\dfrac{dP_\theta}{d\mu}=p(x,\ \theta)$ (μ 为 σ 有限测度)，若 $g(\theta)$ 的估计 $\delta(x)$ 在平方损失下风险有限，$Q(\theta)>0$ 满足引理 2.2 的条件，且存在 $Q_1(\theta)>0$ 对 $\theta_0\leqslant t<b$ $(a<t\leqslant\theta_0)$ 有

$$\left|\int_{\theta_0}^{t}(\delta(x)-g(\theta))Q(\theta)Q_1(\theta)p(x,\ \theta)d\theta\right|$$
$$\leqslant KQ(t)Q_1^{1/2}(t)p(x,\ t),$$
则 $\delta(x)$ 在 $\theta_0\leqslant\theta<b$ $(a<\theta\leqslant\theta_0)$是几乎容许估计. 若进一步假设定理 2.1 的注成立,则 δ 为容许估计.

证明如同定理 2.1, 利用引理 2.2 即得证.

定理2.6 若 X 遵从单参数指数族分布, 参数的自然区域 $\Omega=(a,\ b)$ 为开区间 (a 可为 $-\infty$, b 可为 $+\infty$), $\Theta=(\theta_0,\ b)$ (或 $(a,\ \theta_0)$), $a<\theta_0<b$. 若存在 $\lambda>0$ 及 k, 使得

$$\left|\int_{\theta_0}^{t}\beta^{-\lambda}(\theta)e^{-k\lambda\theta}\,d\theta\right|\to\infty \quad (\text{当 } t\to b(a)).$$

令 $T_1(x)=(T(x)+\lambda k)/(1+\lambda)$, 则

$$\hat{g}(x)=\begin{cases}T_1(x)+\beta^{1+\lambda}(\theta_0)e^{\theta_0(T(x)+\lambda k)}(1+\lambda)^{-1} \\ \quad\times\left(\int_{\theta_0}^{b(a)}\beta^{1+\lambda}(\theta)e^{\theta(T(x)+\lambda k)}\,d\theta\right)^{-1}, \text{当 } t_1<T_1(x)<t_2; \\ t_1, \quad \text{当} \quad T_1(x)\leqslant t_1; \\ t_2, \quad \text{当} \quad T_1(x)\geqslant t_2.\end{cases}$$

是 $g(\theta)=E_\theta T(x)$ 在 Θ 上的容许估计. 此处

$$t_1=\sup\{t;\ \mu(-\infty<T(x)<t)=0\};$$
$$t_2=\inf\{t;\ \mu(t<T(x)<+\infty)=0\}.$$

证 当 $t_1<T_1(x)<t_2$ 时,若 $b=\infty$,

$$\begin{aligned}&\lim_{\theta\to\infty}\theta^2\beta^{1+\lambda}(\theta)e^{\theta(T(x)+\lambda k)} \\ &\quad=\lim_{\theta\to\infty}\theta^2\left[e^{\theta T_1(x)}\left(\int e^{\theta T(x)}\,d\mu\right)^{-1}\right]^{1+\lambda} \\ &\quad=\lim_{\theta\to\infty}\theta^2\left(\int e^{\theta(T(y)-T_1(x))}\,d\mu(y)\right)^{-1-\lambda} \\ &\quad\leqslant\lim_{\theta\to\infty}\theta^2\left(\int_{T(y)>T_1(x)+s}e^{\theta s}\,d\mu\right)^{-1-\lambda}=0.\end{aligned}$$

若 $a(b)$有限, 因为假设 Ω 为开区间, $a(b)$不属于 Ω, 故 $\beta(\theta)\to0$ $(\theta\to a(b))$. 综合上面,就有

$$\int_\Theta\beta^{1+\lambda}(\theta)e^{\theta(T(x)+\lambda k)}\,d\theta<\infty; \tag{2.10}$$

$$\lim_{\theta\to a(b)} \beta^{1+\lambda}(\theta)e^{\theta(T(x)+\lambda k)}=0. \tag{2.11}$$

(2.10)式可用来说明定义的 $\hat{g}$ 有意义. 此时对 (θ_0, b) 的情形, 经计算,有

$$\int_0^z (\hat{g}(x)-g(\theta))\beta^{1+\lambda}(\theta)e^{\theta(T(x)+\lambda k)}d\theta$$

$$=\int_{\theta_0}^z \Big[T_1(x)+\beta^{1+\lambda}(\theta_0)e^{\theta_0(T(x)+\lambda k)}(1+\lambda)^{-1}$$

$$\times\left(\int_{\theta_0}^b \beta^{1+\lambda}(\theta)e^{\theta(T(x)+\lambda k)}d\theta\right)^{-1}+\beta'(\theta)\beta^{-1}(\theta)\Big]$$

$$\times\beta^{1+\lambda}(\theta)e^{\theta(T(x)+\lambda k)}d\theta$$

$$=(1+\lambda)^{-1}\Big[\beta^{1+\lambda}(z)e^{z(T(x)+\lambda k)}-\beta^{1+\lambda}(\theta_0)e^{\theta_0(T(x)+\lambda k)}$$

$$+\beta^{1+\lambda}(\theta_0)e^{\theta_0(T(x)+\lambda k)}\int_{\theta_0}^z \beta^{1+\lambda}(\theta)e^{\theta(T(x)+\lambda k)}d\theta\Big]$$

$$\times\left(\int_{\theta_0}^b \beta^{1+\lambda}(\theta)e^{\theta(T(x)+\lambda k)}d\theta\right)^{-1}$$

$$=(1+\lambda)^{-1}\Big[\beta^{1+\lambda}(z)e^{z(T(x)+\lambda k)}-\beta^{1+\lambda}(\theta_0)e^{\theta_0(T(x)+\lambda k)}$$

$$\times\int_z^b \beta^{1+\lambda}(\theta)e^{\theta(T(x)+\lambda k)}d\theta\left(\int_{\theta_0}^b \beta^{1+\lambda}(\theta)e^{\theta(T(x)+\lambda k)}d\theta\right)^{-1}\Big]$$

$$=(1+\lambda)^{-1}\int_z^b \beta^{1+\lambda}(\theta)e^{\theta(T(x)+\lambda k)}[f(z)-f(\theta_0)]d\theta. \tag{2.12}$$

其中

$$f(z)\triangleq\beta^{1+\lambda}(z)e^{z(T(x)+\lambda k)}\left(\int_z^b \beta^{1+\lambda}(\theta)e^{\theta(T(x)+\lambda k)}d\theta\right)^{-1}; \tag{2.13}$$

$$f'(z)=f(z)[T(x)+\lambda k-(1-\lambda)g(z)+f(z)]$$

由于 $g'(\theta)=\mathrm{Var}\,T(x)>0$, 所以 $g(\theta)$ 为增函数,从而

$$(1+\lambda)g(z)<(1+\lambda)\int_z^b g(\theta)\beta^{1+\lambda}(\theta)e^{\theta(T(x)+\lambda k)}d\theta$$

$$\left(\int_z^b \beta^{1+\lambda}(\theta)e^{\theta(T(x)+\lambda k)}d\theta\right)^{-1}=T(x)+\lambda k+f(z).$$

由此可知, $f'(z)>0$, $f(z)-f(\theta_0)\geqslant 0$ (当 $b>z>\theta_0$). 从(2, 12)知:

$$0\leqslant\int_{\theta_0}^{z}(\hat{g}-g(\theta))\beta^{1+\lambda}(\theta)e^{\theta(T(x)+\lambda k)}d\theta$$

$$\leqslant(1+\lambda)^{-1}\beta^{1+\lambda}(z)e^{z(T(x)+\lambda k)}, \tag{2.14}$$

当 $T_1(x)\geqslant t_2$, $\hat{g}=t_2$. 由 t_1、t_2 的定义知, $t_1\leqslant g(\theta)\leqslant t_2$, 故

$$0\leqslant\int_{\theta_0}^{z}(t_2-g(\theta))\beta^{1+\lambda}(\theta)e^{\theta(T(x)+\lambda k)}d\theta$$

$$=(1+\lambda)^{-1}\beta^{1+\lambda}(\theta)e^{\theta(T(x)+\lambda k)}\Big|_{\theta_0}^{z}$$

$$+\int_{\theta_0}^{z}\left(t_2-\frac{T(x)+\lambda k}{1+\lambda}\right)\beta^{1+\lambda}(\theta)e^{\theta(T(x)+\lambda k)}d\theta$$

$$\leqslant(1+\lambda)^{-1}\beta^{1+\lambda}(z)e^{z(T(x)+\lambda k)}, \tag{2.15}$$

当 $T_1(x)\leqslant t_1$ 时, $\hat{g}=t_1$, 同理可知

$$\left|\int_{\theta_0}^{z}(t_1-g(\theta))\beta^{1+\lambda}(\theta)e^{\theta(T(x)+\lambda k)}d\theta\right|$$

$$\leqslant(1+\lambda)^{-1}\beta^{1+\lambda}(z)e^{z(T(x)+\lambda k)}. \tag{2.16}$$

我们取 $Q(\theta)=\beta^{\lambda}(\theta)e^{\lambda k\theta}$, $Q_1(\theta)=1$, (2.14)～(2.16)式实际上满足定理 2.5 中不等式的要求,由此本定理得证.

例 2.4 设 $X\sim N(\theta, 1)$, $\Theta=(0, \infty)$, 则

$$\delta(x)=x+e^{-x^2/2}\Big/\int_{-\infty}^{x}e^{-t^2/2}dt\triangleq x+V(x)$$

是 θ 在 Θ 上关于平方损失下的容许 minimax 估计.

证 由于正态分布属于指数族, $\beta(\theta)=e^{-\theta^2/2}$, 令 $\lambda=0$,

$$\int_{0}^{x}d\theta\to\infty \quad (\text{当 } x\to\infty).$$

由定理 2.6 可得 $\delta(x)$ 为 $(0, \infty)$ 上 θ 的容许估计. 再注意: $V'(x)=-V(x)[x+V(x)]$, 利用分部积分, 得

$$E_\theta(X-\theta)V(x)=(2\pi)^{-\frac{1}{2}}\int_{-\infty}^{\infty}(x-\theta)V(x)e^{-(x-\theta)^2/2}d\theta$$

$$=(2\pi)^{-\frac{1}{2}}\int_{-\infty}^{\infty}V'(x)e^{-(x-\theta)^2/2}dx$$

$$=-E_\theta[V(X)(X+V(X))];$$

$$R(\theta, \delta)=E_\theta(\theta-\delta(x))^2$$

$$=1+2E_\theta V(X)(X-\theta)+E_\theta V^2(X)$$

$$=1-\theta E_\theta V(X)\leqslant1 \quad (\text{当 } \theta\in(0, \infty)).$$

利用例 2.1 的结论即证得 $\delta(x)$ 的 minimax 性.

§3 单个位置参数 Pitman 估计的容许性

对于转移分布族

$$P(x-\theta\mathbb{1}),\ \theta\in R^1,\ x\in R^k,\ \mathbb{1}=(1,\ \cdots,\ 1)'_{k\times 1},$$

Pitman 在 1939 年曾从 Bayes 观点提出了一个有名的 θ 的估计 $\hat{\theta}=X_1+E_0(X_1|Y)$, 这里 $Y=(X_2-X_1,\ \cdots,\ X_k-X_1)$. 在前一章我们已证明了此估计是最优同变估计, 本节将讨论此估计的容许性, 介绍 Stein 的工作[26, iii].

为了讨论方便起见, 记 $X_1|Y\sim F(X-\theta|Y)$, $Y\sim\nu(y)$. 易知 Y 是最大不变量, 它的分布与参数无关. 若 $E_0(X_1|Y)\neq 0$, 取 $X_i'=X_i-E_0(X_1|Y)$ $(i=1,\ 2,\ \cdots,\ k)$, 则 $E_0(X_1'|Y)=0$. 因此, 不失一般性, 假设 $E_0(X_1|Y)=0$.

定理 3.1 若 $E_0(X_1|Y)=0$;

$$E\{E_0(X_1^2|Y)\}^{3/2}<\infty.$$

则在平方损失下, x_1 是 θ 的几乎容许估计. 若进一步假设 $dF(x|y)=p(x|y)dx$, 则 x_1 是 θ 的容许 minimax 估计.

证 我们采用定理 1.4 的办法来证明. 如果 x_1 不是 θ 的几乎容许估计, 则存在 θ 的另一估计 $\varphi(X,\ Y)$, 使得当 $-\infty<\theta<\infty$ 时, 有

$$E_\theta(\varphi(X,\ Y)-\theta)^2\leqslant E_\theta(X_1-\theta)^2=\int d\nu\int x^2dF(x|y),\quad (3.1)$$

且使 (3.1) 严格不等号成立的所有 θ 的集合 s, 有 $L(s)>0$. 因此存在 $\varepsilon>0$ 及有界集合 E_ε, 使 $L(E_\varepsilon)>0$, 并且当 $\theta\in E_\varepsilon$ 时, 有

$$E_\theta(\varphi(X,\ Y)-\theta)^2\leqslant E_\theta(X_1-\theta)^2-\varepsilon.\qquad (3.2)$$

取 θ 的先验分布为 Cauchy 分布 $Q(\theta\sigma^{-1})$, 有密度函数

$$\sigma^{-1}q(\theta\sigma^{-1})=\sigma^{-1}\pi^{-1}(1+\theta^2\sigma^{-2})^{-1}\quad(\sigma>0,\ \theta\in R^1).$$

由 (3.1) 与 (3.2) 式可得

$$R(\sigma,\ \varphi)\triangleq\int E_\theta(\varphi(X,\ Y)-\theta)^2\sigma^{-1}q(\theta\sigma^{-1})d\theta$$
$$\leqslant E_\theta(X_1-\theta)^2-\varepsilon\int_{E_\varepsilon}\sigma^{-1}q(\theta\sigma^{-1})d\theta$$
$$\leqslant E_0X_1^2-\varepsilon kL(E_\varepsilon)/\sigma\quad(\sigma>1).\tag{3.3}$$

此处常数 $k=\pi^{-1}\inf\limits_{\theta\in E_\varepsilon}(1+\theta^2)^{-1}>0$.

我们定义 $\sigma^{-1}f(\sigma)=E_0X_1^2-\inf\limits_{\varphi}R(\sigma,\ \varphi)$，如果能够证明 $f(\sigma)\to 0$(当 $\sigma\to\infty$ 时)，则与(3.3)式产生矛盾，否定了反证的假设，定理就能得证.

首先易证 $\inf R(\sigma,\ \varphi)$ 在 $\varphi=E(\theta|X_1,\ Y)$ 达到. 这样

$$\sigma^{-1}f(\sigma)=E_\sigma(X_1-\theta)^2-E_\sigma[E(\theta|X_1,\ Y)-\theta]^2$$
$$=E_\sigma[X_1^2-2\theta X_1+(E(\theta|X_1,\ Y))^2]$$
$$=E_\sigma[X_1-E(\theta|X_1,\ Y)]^2.\tag{3.4}$$

这里的 E_σ 是对 $(\mathscr{X}\times\Theta,\ \mathscr{B}_x\times\mathscr{B}_\Theta)$ 上的概率测度(为 $dP(x-\theta|)\times\sigma^{-1}q(\theta\sigma^{-1})d\theta$)求数学期望. 为了进一步弄清 $f(\sigma)$ 的明确表达式，先需证明对任一 $B\in\mathscr{B}_\Theta$，有

$$P(B|x,\ y)=\int_B q(\theta\sigma^{-1})d_\theta F(x-\theta|y)$$
$$\times\left(\int_{-\infty}^{\infty}q(\theta\sigma^{-1})d_\theta F(x-\theta|y)\right)^{-1}.\tag{3.5}$$

为证(3.5)，先给出 $X_1|Y$ 的绝对无条件分布为

$$P_\sigma(x,\ y)=\sigma^{-1}\int_{-\infty}^{\infty}q(\theta\sigma^{-1})F(x-\theta|y)d\theta$$
$$=\int_{-\infty}^{\infty}F(x-\theta|y)d_\theta Q(\theta\sigma^{-1})$$
$$=-\int_{-\infty}^{\infty}Q(\theta\sigma^{-1})d_\theta F(x-\theta|y)$$
$$=\int_{-\infty}^{\infty}Q\left(\frac{x-\theta}{\sigma}\right)d_\theta F(\theta|y).$$

其中第三个等式是对 $R-S$ 积分实行分部积分. 在最后第二式中令 $x-\theta=z$，再将 z 换为 θ，即得最后一式. 将上式微分，得

$$\frac{\partial}{\partial x}P_\theta(x,\ y)=\int_{-\infty}^{\infty}\sigma^{-1}q\left(\frac{x-\theta}{\sigma}\right)d_\theta F(\theta|y)$$

$$=-\int_{-\infty}^{\infty}\sigma^{-1}q(\theta\sigma^{-1})d_\theta F(x-\theta|y).$$

于是对任意 $A\in\mathscr{B}_x$, $B\in\mathscr{B}_\theta$ 及任意实数 t, 有

$$\int_A\int_{-\infty}^{t}\int_{-\infty}^{\infty}\sigma^{-1}I_B(\theta)q(\theta\sigma^{-1})d_\theta F(x-\theta|y)dx\,d\nu(y)$$

$$=-\int_A\int_{-\infty}^{t}\int_{-\infty}^{\infty}\sigma^{-1}I_B(x-\theta)q\left(\frac{x-\theta}{\sigma}\right)d_\theta F(\theta|y)dx\,d\nu(y)$$

$$=-\int_A\int_{-\infty}^{\infty}\int_{-\infty}^{t}\sigma^{-1}I_B(x-\theta)q\left(\frac{x-\theta}{\sigma}\right)dx\,d_\theta F(\theta|y)d\nu(y)$$

$$=-\int_A\int_{-\infty}^{\infty}\int_{-\infty}^{t-\theta}\sigma^{-1}I_B(z)q(z\sigma^{-1})dz\,d_\theta F(\theta|y)d\nu(y)$$

$$=\int_A\int_{-\infty}^{\infty}\int_{-\infty}^{\xi}\sigma^{-1}I_B(z)q(z\sigma^{-1})dz\,d_\xi F(t-\xi|y)d\nu(y)$$

$$=-\int_{-\infty}^{\infty}\int_A F(t-\theta|y)I_B(\theta)q(\theta\sigma^{-1})d\nu(y)d\theta.$$

若记(3.5)右边为 $K(B;\ x,\ y)$, 则上一等式说明

$$\iint_C K(B;\ x,\ y)\frac{\partial}{\partial x}P_\sigma(x,\ y)dx\,d\nu(y)$$

$$=P((X_1,\ Y)\in C,\ \theta\in B)$$

对于任意 $C=A\times(-\infty,\ t)$ 皆成立. 结合上面等式及条件概率的定义, 即可推得(3.5)成立, 再由条件概率与条件期望的关系式知

$$E(\theta|X_1,\ Y)=\int_{-\infty}^{\infty}\theta q(\theta\sigma^{-1})d_\theta F(x-\theta|y)$$

$$\times\left(\int_{-\infty}^{\infty}q(\theta\sigma^{-1})d_\theta F(x-\theta|y)\right)^{-1}. \qquad (3.6)$$

由(3.4)、(3.6)式导出 $f(\sigma)$ 的表达式如下:

$$f(\sigma)=\int_{-\infty}^{\infty}q(\theta\sigma^{-1})d\theta\int d\nu$$

$$\times\int_{-\infty}^{\infty}\left\{x-\frac{\int_{-\infty}^{\infty}\theta' q(\theta'\sigma^{-1})d_{\theta'}F(x-\theta'|y)}{\int_{-\infty}^{\infty}q(\theta'\sigma^{-1})d_{\theta'}F(x-\theta'|y)}\right\}^2 d_xF(x-\theta|y)$$

$$
\begin{aligned}
&=\int d\nu\int_{-\infty}^{\infty} q(\theta\sigma^{-1})d\theta \\
&\quad\times\int_{-\infty}^{\infty}\left\{\frac{\int_{-\infty}^{\infty}\eta q\left(\frac{x-\eta}{\sigma}\right)d_{\eta}F(\eta|y)}{\int_{-\infty}^{\infty}q\left(\frac{x-\eta}{\sigma}\right)d_{\eta}F(\eta|y)}\right\}^{2}d_{x}F(x-\theta|y) \\
&=\int d\nu\int_{-\infty}^{\infty} q(\theta\sigma^{-1})d\theta \\
&\quad\times\int_{-\infty}^{\infty}\left\{\frac{\int_{-\infty}^{\infty}\eta q\left(\frac{\theta+z-\eta}{\sigma}\right)d_{\eta}F(\eta|y)}{\int_{-\infty}^{\infty}q\left(\frac{\theta+z-\eta}{\sigma}\right)d_{\eta}F(\eta|y)}\right\}^{2}d_{z}F(z|y) \\
&=\int d\nu\int_{-\infty}^{\infty} d_{z}F(z|y) \\
&\quad\times\int_{-\infty}^{\infty}\left\{\frac{\int_{-\infty}^{\infty}\eta q\left(\frac{x-\eta}{\sigma}\right)d_{\eta}F(\eta|y)}{\int_{-\infty}^{\infty}q\left(\frac{x-\eta}{\sigma}\right)d_{\eta}F(\eta|y)}\right\}^{2}q\left(\frac{x-z}{\sigma}\right)dx \\
&=\int d\nu\int_{-\infty}^{\infty}\frac{\left\{\int_{-\infty}^{\infty}\eta q\left(\frac{x-\eta}{\sigma}\right)d_{\eta}F(\eta|y)\right\}^{2}}{\int_{-\infty}^{\infty}q\left(\frac{x-\eta}{\sigma}\right)d_{\eta}F(\eta|y)}dx. \qquad (3.7)
\end{aligned}
$$

考察(3.7)右端的被积函数，记

$$
\begin{aligned}
\Phi(\sigma,\ F)=&\int_{-\infty}^{\infty}\left\{\int_{-\infty}^{\infty}\eta q\left(\frac{x-\eta}{\sigma}\right)d_{\eta}F(\eta)\right\}^{2} \\
&\times\left\{\int_{-\infty}^{\infty}q\left(\frac{x-\eta}{\sigma}\right)d_{\eta}F(\eta)\right\}^{-1}dx
\end{aligned}
$$

对 $\Phi(\sigma, F)$ 的被积函数分子作 Schwarz 不等式，得

$$
\begin{aligned}
\Phi(\sigma,\ F)&\leqslant\int_{-\infty}^{\infty}\left\{\int\eta^{2}q\left(\frac{x-\eta}{\sigma}\right)d_{\eta}F(\eta)\right\}dx \\
&=\sigma\int\eta^{2}dF(\eta)\triangleq\sigma\lambda. \qquad (3.8)
\end{aligned}
$$

令 U_{λ} 为满足 $\int_{-\infty}^{\infty}\eta\,dF(\eta)=0$, $\int_{-\infty}^{\infty}\eta^{2}dF(\eta)=\lambda$ 的所有概率分布的集合，并令

$$
\psi(\lambda,\ \sigma)=\sup_{F\in U_{\lambda}}\Phi(\sigma,\ F).
$$

由(3.8)显而易见有

$$\psi(\lambda,\ \sigma)=\sigma^3\psi(\lambda/\sigma^2,\ 1)\leqslant\sigma\lambda. \tag{3.9}$$

当 $\lambda\leqslant 1/2,\ F\in U_\lambda$ 时，有

$$\begin{aligned}\int q(x-\eta)dF(\eta)&\geqslant F\{(-1,\ 1)\}\inf_{|\eta|<1}q(x-\eta)\\&\geqslant\int_{|\eta|<1}(3\pi(1+x^2))^{-1}dF(\eta)\\&\geqslant(1-\lambda)(3\pi(1+x^2))^{-1}\geqslant(6\pi(1+x^2))^{-1}.\end{aligned}$$

使用 Chebyshev 不等式得

$$\begin{aligned}&\left(\int_{-\infty}^{\infty}\eta q(x-\eta)dF(\eta)\right)^2\\&\quad=\left(\int_{-\infty}^{\infty}\eta[q(x-y)-q(x)]dF(\eta)\right)^2\\&\quad\leqslant\int_{-\infty}^{\infty}\eta^2dF(\eta)\cdot\int_{-\infty}^{\infty}[q(x-\eta)-q(x)]^2dF(\eta).\end{aligned}$$

将上面两个不等式代入 $\Phi(1,\ F)$ 式中，即得

$$\begin{aligned}\Phi(1,\ F)&\leqslant 6\pi\int_{-\infty}^{\infty}(1+x^2)dx\int_{-\infty}^{\infty}\eta^2dF(\eta)\\&\quad\times\int_{-\infty}^{\infty}[q(x-\eta)-q(x)]^2dF(\eta)\\&=\frac{6\lambda}{\pi}\int_{-\infty}^{\infty}dF(\eta)\int_{-\infty}^{\infty}[(1+(x-\eta)^2)^{-1}-(1+x^2)^{-1}]^2(1+x^2)dx\\&=6\lambda\pi^{-1}\int_{-\infty}^{\infty}dF(\eta)\int_{-\infty}^{\infty}\left\{\frac{1+x^2}{(1+(x-\eta)^2)^2}\right.\\&\quad\left.-\frac{2}{1+(x-\eta)^2}+\frac{1}{1+x^2}\right\}dx\\&=6\lambda\pi^{-1}\int_{-\infty}^{\infty}\left\{\int\left[\frac{1+(\eta+z)^2}{(1+z^2)^2}-\frac{1}{1+z^2}\right]dz\right\}dF(\eta)\\&=6\lambda\pi^{-1}\int_{-\infty}^{\infty}dF(\eta)\int_{-\infty}^{\infty}\frac{\eta^2+2\eta z}{(1+z^2)^2}dz=c\lambda^2,\end{aligned}\tag{3.10}$$

此处

$$c=6\pi^{-1}\int_{-\infty}^{\infty}(1+x^2)^{-2}dx.$$

由(3.9)与(3.10)式即知

$$\psi(\lambda,\ \sigma)\leqslant\begin{cases}c\lambda^2\sigma^{-1}, & \text{当 } \lambda\sigma^{-2}\leqslant 1/2;\\ \sigma\lambda, & \text{其它。}\end{cases}$$

这样就有

$$f(\sigma)=\int d\nu\int_{-\infty}^{\infty}\frac{\left\{\int_{-\infty}^{\infty}\eta q\left(\frac{x-\eta}{\sigma}\right)dF(\eta|y)\right\}^2}{\int_{-\infty}^{\infty}q\left(\frac{x-\eta}{\sigma}\right)dF(\eta|y)}dx$$

$$\leqslant c\sigma^{-1}\int_{0\leqslant\lambda(y)\leqslant\sigma^2/2}\lambda^2(y)d\nu(y)+\sigma\int_{\sigma^2/2<\lambda<\infty}\lambda(y)d\nu(y)$$

$$\leqslant c\sqrt{\varepsilon/2}\int_{0\leqslant\lambda\leqslant\varepsilon\sigma^2/2}\lambda^{3/2}(y)d\nu(y)$$

$$+c2^{-\frac{1}{2}}\int_{\varepsilon\sigma^2/2<\lambda\leqslant\sigma^2/2}\lambda^{3/2}(y)d\nu(y)$$

$$+\sqrt{2}\int_{\sigma^2/2<\lambda<\infty}\lambda^{3/2}(y)d\nu(y)$$

$$\to c\sqrt{\varepsilon/2}\int_0^{\infty}\lambda^{3/2}(y)d\nu\quad(\sigma\to\infty).$$

对任意 $0<\varepsilon<1$ 成立. 由于 ε 的任意性,得 $f(\sigma)\to 0(\sigma\to\infty)$. 这与(3.3)式矛盾. 于是定理第一部分得证.

下面证第二部分,令

$$S=\{(x_1,y):\ \varphi(x_1,y)\neq x_1\},\ S_y=\{x_1:\ x_1\neq\varphi(x,y)\},$$

若 S 的 $L\times\nu$ 测度为正,则

$$\int_{-\infty}^{\infty}d\theta\int d\nu\int_{S_y}p(x-\theta|y)dx=\int d\nu\int_{S_y}dx\int p(x-\theta|y)d\theta$$

$$=\int d\nu\int_{S_y}dx=L\times\nu(S)>0.$$

因此存在一个 L 正测度集合 T,使得

$$\int d\nu\int_{S_y}p(x-\theta|y)dx>0,\quad 当\ \theta\in T.$$

定义 $\varphi_0(x_1,y)=(x_1+\varphi(x_1,y))/2$,则易知当 $(x_1,y)\in S$ 时,有

$$\varphi_0(x_1,y)-\theta)^2<[(x_1-\theta)^2+(\varphi(x_1,y)-\theta)^2]/2. \quad (3.11)$$

而当 $(x_1,y)\overline{\in}S$ 时,上述不等式取等号,注意(3.11)式,即得

$$E_\theta[\varphi_0(X_1,Y)-\theta]^2<[E_\theta(\varphi(X_1,Y)-\theta)^2$$

$$+E_\theta(X_1-\theta)^2]/2\leqslant E_\theta(X_1-\theta)^2\quad(当\ \theta\in T\ 时).$$

这与 X_1 几乎容许相矛盾. 故 $L\times\nu(S)=0$. 这说明 x_1 是 θ 的容

许估计，又因它的风险为常数，故也是 minimax 估计，至此定理证完.

定理 3.2 若 $X=(X_1, X_2, \cdots, X_k)'$ 有概率密度 $p(x-\mathbb{1}\theta)$, $\theta\in R^1$, 且若

$$\int_{R^k}\left\{\frac{\int\theta^2 p(x-\mathbb{1}\theta)d\theta}{\int p(x-\mathbb{1}\theta)d\theta}-\left(\frac{\int\theta p(x-\mathbb{1}\theta)d\theta}{\int p(x-\mathbb{1}\theta)d\theta}\right)^2\right\}^{3/2}p(x)dx<\infty. \tag{3.12}$$

则 Pitman 估计 $\hat{\theta}=\int_{-\infty}^{\infty}\theta p(x-\mathbb{1}\theta)d\theta\Big/\int_{-\infty}^{\infty}p(x-\mathbb{1}\theta)d\theta$ 是 θ 的容许 minimax 估计(在平方损失下).

证 令 $Y=(X_2-X_1, \cdots, X_k-X_1)'$, 则 $(X_1, Y')'$ 的密度函数为 $p(x_1-\theta, x_1+y_1-\theta, \cdots, y_{k-1}+x_1-\theta)$, 可见 Y 的密度函数为

$$\begin{aligned}g(y)&=\int_{-\infty}^{\infty}p(x_1-\theta, y_1+x_1-\theta, \cdots, y_{k-1}+x_1-\theta)dx_1\\&=\int_{-\infty}^{\infty}p(\eta, y_1+\eta, \cdots, y_{k-1}+\eta)d\eta\\&=\int_{-\infty}^{\infty}p(x_1-\theta', x_2-\theta', \cdots, x_k-\theta')d\theta'\\&=\int p(x-\theta\mathbb{1})d\theta.\end{aligned}$$

其中作变量替换 $\eta=x-\theta$ 以及 $\eta=x-\theta'$.

记 $Z=X_1-E_0(X_1|Y)$, 显然, $E_0(Z|Y)=0$. 而

$$\begin{aligned}E_0(X_1^i|Y)&=(g(y))^{-1}\int_{-\infty}^{\infty}x^i p(x, y_1+x, \cdots, y_{k-1}+x)dx\\&=(g(y))^{-1}\int_{-\infty}^{\infty}(x-\theta)^i p(x-\mathbb{1}\theta)d\theta\quad(i=1, 2).\end{aligned}\tag{3.13}$$

所以由(3.13)易知

$$\begin{aligned}E_0(Z^2|Y)&=E_0(X_1^2|Y)-[E_0(X_1|Y)]^2\\&=\frac{\int_{-\infty}^{\infty}\theta^2 p(x-\theta\mathbb{1})d\theta}{\int_{-\infty}^{\infty}p(x-\theta\mathbb{1})d\theta}-\left(\frac{\int_{-\infty}^{\infty}\theta p(x-\theta\mathbb{1})d\theta}{\int_{-\infty}^{\infty}p(x-\theta\mathbb{1})d\theta}\right)^2.\end{aligned}$$

故由题设知，定理 3.1 的假设成立，因此 θ 的 Pitman 估计

$$z=\int_{-\infty}^{\infty}\theta p(x-\theta\mathbb{1})d\theta\Big/\int_{-\infty}^{\infty}p(x-\theta\mathbb{1})d\theta$$

是容许 minimax 估计．定理得证．

定理 3.4 设 $X=(X_1, X_2, \cdots, X_k)'$ 具有概率密度

$$\theta^{-k}p(x/\theta)\quad (0<\theta<\infty),\ x=(x_1, x_2, \cdots, x_k)'.$$

若损失函数为 $L(\theta, d)=(\log\theta-d)^2$ 以及

$$\int_{R^k}p(x)dx\left\{\frac{\int_0^{\infty}\theta^{-k-1}(\log\theta)^2p(x/\theta)d\theta}{\int_0^{\infty}\theta^{-k-1}p(x/\theta)d\theta}\right.$$

$$\left.-\left(\frac{\int_0^{\infty}\theta^{-k-1}\log\theta p(x/\theta)d\theta}{\int_0^{\infty}\theta^{-k-1}p(x/\theta)d\theta}\right)^2\right\}^{3/2}<\infty.$$

则
$$\widehat{\log\theta}=\int_0^{\infty}\theta^{-k-1}\log\theta p(x/\theta)d\theta\Big/\int_0^{\infty}\theta^{-k-1}p(x/\theta)d\theta$$

是 $\log\theta$ 的容许 minimax 估计．

证 把刻度参数化为位置参数来处理，令 $Y=(X_2/X_1, \cdots, X_k/X_1)$，以 $\log|X_1|$ 代替过去位置参数的 X_1，利用定理 3.1，即可证得定理．证明细节留作练习．

注 1 定理 3.1～3.3 的主要条件比较复杂，不便验证．现在给出方便的充分条件：若存在同变估计 $\varphi(X)$，使 $E_0(|\varphi(X)|^3)<\infty$（对于刻度参数改为 $E_1|\varphi(X)|^3<\infty$），则定理的结论仍然成立．

证 因 φ 是同变估计 $\varphi(x-x_1\mathbb{1})=\varphi(x)-x_1$，故

$$E_0(\varphi(X)|X_2-X_1, \cdots, X_k-X_1)-E_0(X_1|X_2-X_1, \cdots, X_k-X_1)=\varphi(x-x_1\mathbb{1})=\varphi(x)-x_1$$

$$\Rightarrow \hat{\theta}=\varphi(X)-E_0(\varphi(X)|x_2-x_1, \cdots, x_k-x_1)$$
$$=x_1-E_0(x_1|x_2-x_1, \cdots, x_k-x_1).$$

由于$|a^3|$是a的严凸函数，$|a+b|^3\leqslant 4(|a^3|+|b^3|)$，故而

$$E_0|\hat{\theta}|^3\leqslant 4[E_0|\varphi(X)|^3+E_0|E_0(\varphi(X)|X_2-X_1,\ \cdots,\ X_k-X_1)|^3]\leqslant 8E_0|\varphi(X)|^3<\infty.$$

此外 $E_0\{E_0(\hat{\theta}^2|X_2-X_1,\ \cdots,\ X_k-X_1)\}^{3/2}\leqslant E_0|\hat{\theta}|^3<\infty$.

因此定理 3.1 的条件成立. 对刻度参数问题，可同样论证.

注 2 如果定理 3.1 不假设 $P(x|y)$ 有概率密度，则最优同变估计不一定是容许的. Blackwell[3] 曾给过反例.

注 3 是否存在比定理 3.1 更弱的条件而使定理的结论仍成立，Perng 给过反例[24]，说明除非另加条件，否则，3/2 的幂次是不能降低的.

例 3.1 i) 若 $x_1,\ x_2,\ \cdots,\ x_n$ 为 iid $(n>1)$，皆具有 $\left[\theta-\frac{1}{2},\ \theta+\frac{1}{2}\right]$ 上的均匀分布，它的顺序统计量为 $x_{(1)}\leqslant x_{(2)}\leqslant\cdots\leqslant x_{(n)}$，则 $\hat{\theta}_n=(x_{(1)}+x_{(n)})/2$ 为 θ 的容许估计(平方损失).

易见分布属于转移参数族，$E_0|\hat{\theta}_n|^3<\infty$. 故定理 3.1 成立. 其 Pitman 估计为

$$\hat{\theta}_n=\int_{x_{(n)}-1/2}^{x_{(1)}+1/2}\theta\,d\theta\Big/\int_{x_{(n)}-1/2}^{x_{(1)}+1/2}d\theta=[x_{(1)}+x_{(n)}]/2$$

是 θ 的容许 minimax 估计.

ii) 若 $x_1,\ x_2,\ \cdots,\ x_n$ $(n\geqslant 7)$ 为 iid，具有 Cauchy 分布密度为 $\pi^{-1}(1+(x-\theta)^2)^{-1}$，则

$$\hat{\theta}_n=\int_{-\infty}^{\infty}\left[\prod_1^n\frac{\theta}{1+(x_i-\theta)^2}\right]d\theta\Big/\int_{-\infty}^{\infty}\left[\prod_1^n(1+(x_i-\theta)^{-2})\right]d\theta$$

是 θ 的容许 minimax 估计.

由于 Cauchy 分布为转移参数族，当 $n\geqslant 7$ 时，其样本中的位数(当 n 为奇数时为 $x_{\left(\frac{n+1}{2}\right)}$；当 n 为偶数时为 $[x_{\left(\frac{n}{2}\right)}+x_{\left(\frac{n}{2}+1\right)}]/2$) 的三阶矩存在有限. 故满足注 1 的条件. 结论得证.

iii) 令 $x_1,\ x_2,\ \cdots,\ x_n$ 为 iid, $\sim N(0,\ \sigma^2)$. 因为 $\log\sigma$ 的一个同变估计 $\log\sqrt{\sum_1^n x_i^2/n}$ 的三阶矩存在，故应用注 1 知

$$\widehat{\log\sigma}=\int_0^\infty \sigma^{-n-1}(\log\sigma)\exp\left[-\sum_1^n x_i^2\Big/2\sigma^2\right]d\sigma\Big/$$

$$\int_0^\infty \sigma^{-n-1}\exp\left[-\sum_1^n x_i^2\Big/2\sigma^2\right]d\sigma$$

$$=\int_0^\infty \frac{1}{2}y^{n/2-1}(\log y)e^{-y\sum_1^n x_i^2/2}dy\Big/\int_0^\infty y^{n/2-1}e^{-y\sum_1^n x_i^2/2}dy$$

$$=\frac{1}{2}\log\frac{1}{2}\sum_{i=1}^n x_i^2-\Gamma'\left(\frac{n}{2}\right)\Big/2\Gamma\left(\frac{n}{2}\right)$$

是 $\log\sigma$ 关于损失函数 $(\log\sigma-d)^2$ 的容许 minimax 估计.

§4 多维正态分布参数估计的容许性问题

多维正态分布在实际应用中有它特殊的作用，也是数理统计中研究比较多的，自然是我们注意的重点. 对它的均值及协方差阵的估计，历来采用样本均值及样本协方差阵，不同的估计只是协方差阵乘上不同的常数因子，原先人们都不怀疑这种估计的优良性. 但是随着容许性理论的发展，情况就有了变化. James 和 Stein 在 50～60 年代举例说明了在平方损失下，三维以上的正态分布的样本均值不是均值的容许估计. 随后又举例说明了依赖于样本方差的线性函数也不是一维正态分布的方差的容许估计，这就进一步揭露了多维正态分布的内在联系，引起近二十年来对这个问题的广泛研究. 本节将介绍这方面的某些研究结果，分下列四个方面阐述之.

一、方差估计的改进

假设 $X\sim N(\mu,\sigma^2)$, $\mu\in R^1$, $\sigma^2>0$ 皆未知，$x_1, x_2, \cdots, x_n$ 是从 X 的简单抽样，我们的目的是估计 σ^a, $a>0$ 已知 ($a=1$ 即估计 σ; $a=2$ 即估计 σ^2). 损失函数为 $(\sigma^a-d)^2/\sigma^{2a}=(\sigma^{-a}d-1)^2$.

首先考虑在仿射变换下的最优同变估计. 众所周知

$$\bar{x}=n^{-1}\sum_{i=1}^n x_i,\quad s^2=\sum_{i=1}^n(x_i-\bar{x})^2$$

构成(μ, σ^2)的充分统计量，又损失函数为凸损失函数，我们可限制在依赖于充分统计量$(\bar{x}, s^2)$的非随机化估计的范围内来研究我们的问题. 在上一章已证明了对位置刻度参数族，在平方损失下估计位置和刻度参数，在仿射变换群$\mathscr{G}=\left\{\begin{pmatrix}1 & 0\\ e & c\end{pmatrix}: c>0, e\in R^1\right\}$变换下为不变判决问题. 变换为对$\forall g\in\mathscr{G}$, $x=(x_1, \cdots, x_n)'$,

$$gx=cx+e\mathbb{1}, \quad \bar{g}(\mu, \sigma)=(c\mu+e, c\sigma), \quad g^*d=c^a d.$$

此处$\bar{g}$、g^*与g对应，分别属于在参数空间与行动空间上与$\mathscr{G}$同态的变换群$\overline{\mathscr{G}}$与$\mathscr{G}^*$. 依赖于$(\bar{x}, s^2)$的刻度参数估计$\delta(\bar{x}, s^2)$应满足

$$\delta(\bar{x}, s^2)=g^{*-1}(\delta(\bar{x}, s^2))=c^{-a}\delta(c\bar{x}+e, c^2s^2). \qquad (4.1)$$

若取$c=s^{-1}$, $e=-s^{-1}\bar{x}$, 得$\delta(\bar{x}, s^2)=s^a\delta(0, 1)$. 故得到$\sigma^a$的一般同变估计形式$cs^a$（这里$c$是常数）. 由此不难求得$\sigma^a$在平方损失下的最优同变估计为$c_0s^a$，此处

$$c_0=E_{0,1}(s^a)/E_{0,1}(s^{2a})=\Gamma\left(\frac{n-1+a}{2}\right)\left(\Gamma\left(\frac{n-1+2a}{2}\right)\right)^{-1}. \qquad (4.2)$$

它仅依赖于样本方差. 而如在本章§1所述，样本均值也提供少量方差的信息. 从直观上说明了最优同变估计c_0s^a还有可能改进. 现考虑在刻度变换群(即仿射群中$e=0$的子群)的同变估计. 在(4.1)中取$e=0$, $c=\left(\sum_{i=1}^{n}x_i^2\right)^{-1/2}\triangleq Q^{-1/2}$. 则得估计

$$d(\bar{x}, s^2)=Q^{\frac{a}{2}}d(\bar{x}Q^{-\frac{1}{2}}, s^2Q^{-1})=Q^{\frac{a}{2}}\psi(s^2Q^{-1}). \qquad (4.3)$$

现仅考虑$d(-\bar{x}, s^2)=d(\bar{x}, s^2)$的情况以及$Q=n\bar{x}^2+s^2$, 故(4.3)式是合理的. 现在要找比$c_0s^a$一致优的(4.3)形的估计. 它是刻度变换群的同变估计. 故风险函数各轨道上是常数，它不依赖σ. 不妨设$\sigma=1$. 此时$(n\bar{x}^2, s^2)$的联合分布密度为

$$(2\pi)^{-\frac{1}{2}}[\exp(-2^{-1}(\sqrt{x}-\sqrt{n}\,\mu)^2)$$
$$+\exp(-2^{-1}(\sqrt{x}+\sqrt{n}\,\mu)^2)]x^{-\frac{1}{2}}f_{n-1}(y)$$

$$=2(2\pi)^{-\frac{1}{2}}\exp\left(-\frac{1}{2}(x+n\mu^2)\right)$$

$$\times\sum_{j=0}^{\infty}x^{j-\frac{1}{2}}(n\mu^2)^j(2j!)^{-1}f_{n-1}(y)$$

$$=\sum_{j=0}^{\infty}f_{1+2j}(x)f_{n-1}(y)P_j(2^{-1}n\mu^2).$$

这里f_k为k个自由度的χ^2分布，$P_j(\lambda)=e^{-\lambda}\lambda^j/j!$，说明$n\bar{x}^2$具有$f_{1+2J}$同样的分布．这里$J\sim$Poisson 分布$\mathscr{P}\left(\frac{n\mu^2}{2}\right)$的随机变量．

在$J=j$条件下，作变换$W=y(x+y)^{-1}$，$Z=x+y$，它分别代表Q和s^2Q^{-1}．由此可知，在$J=j$下，$(Q,\ s^{-2}Q)$的条件概率密度为

$$\left[\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{1+2j}{2}\right)2^{(n+2j)/2}\right]^{-1}z^{(n+2j)/2-1}$$

$$\times\omega^{(n+1)/2-1}(1-\omega)^{(1+2j)/2-1}e^{-\frac{z}{2}}\quad(z>0,\ 0<\omega<1).$$

这说明$s^2Q^{-1}|J=j\sim$Beta 分布$B\left(\frac{n+1}{2},\ \frac{1+2j}{2}\right)$，$Q|J=j\sim\chi^2_{n+2j}$分布，且相互独立．这是正态分布理论中的一个重要结论，据此可得

$$c_b(j)\triangleq E(Q^b|J=j)=2^b\Gamma\left(\frac{n+2j}{2}+b\right)\Big/\Gamma\left(\frac{n+2j}{2}\right).$$

现在计算

$$\begin{aligned}
\bar{R}(\mu)&=R(\mu,\ c_0s^a)-R(\mu,\ d(\bar{x},\ s^2))\\
&=E_{\mu,1}[(c_0s^a-1)^2-(Q^{\frac{a}{2}}\psi(s^2Q^{-1})-1)^2]\\
&=E_{\mu,1}[E\{(c_0s^a-1)^2-(Q^{\frac{a}{2}}\psi(s^2Q^{-1})-1)^2|J=j\}]\\
&=E_{\mu,1}[E\{Q^{\frac{a}{2}}(c_0(s^2Q^{-1})^{\frac{a}{2}}-\psi(s^2Q^{-1}))\\
&\quad\times(Q^{a/2}(c_0(s^2Q^{-1})^{\frac{a}{2}}+\psi(s^2Q^{-1})-2|J=j\}]\\
&=E_{\mu,1}[E\{2c_{\frac{a}{2}}(j)(c_0(s^2Q^{-1})-\psi(s^2Q^{-1}))\\
&\quad\times(c_a(j)c_{a/2}^{-1}(j)2^{-1}(c_0(s^2Q^{-1})^{\frac{a}{2}}+\psi(s^2Q^{-1})-1)|J=j\}]
\end{aligned}\tag{4.4}$$

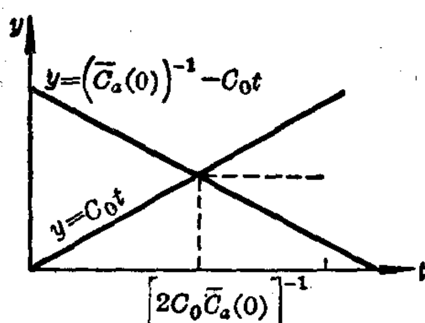
记

$$\bar{c}_a(j)=2^{-1}c_a(j)(c_{\frac{a}{2}}(j))^{-1}$$

$$=2^{a/2-1}\Gamma\left(\frac{n+2j+2a}{2}\right)\left(\Gamma\left(\frac{n+2j+a}{2}\right)\right)^{-1}.$$

它是 j 的增函数且随 $j\to\infty$ 而趋于 ∞，而作为 σ^a 的估计 $\psi>0$ 才合理。故若选择 ψ 使当 $0<\omega$ 时有

$$\begin{gathered} c_a\omega^{\frac{a}{2}}-\psi(\omega)\geqslant 0, \\ (c_0\omega^{\frac{a}{2}}-\psi(\omega))(\bar{c}_a(0)(c_0\omega^{\frac{a}{2}}+\psi(\omega))-1))\geqslant 0. \end{gathered} \tag{4.5}$$



且(4.5)严格不等号成立的集合为正测度，就能使(4.4)式最右端每个 $J=j$ 的条件数学期望为正，因而 $\bar{R}(\mu)>0$ 对一切 $\mu\in R^1$ 成立。

如令 $t=\omega^{\frac{a}{2}}$，$g(t)=\psi(\omega)$，则(4.5)要求可化作

$$\begin{cases} g(t)=c_0t, & \text{当}\quad 0<t<[2c_0\bar{c}_a(0)]^{-1}; \\ c_0t>g(t)>(\bar{c}(0))^{-1}-c_0t, & \text{当}\quad [2c_0\bar{c}_a(0)]^{-1}<t<1. \end{cases} \tag{4.6}$$

现在我们就把上面的论证写为如下定理：

定理 4.1 若 $X\sim N(\mu,\sigma^2)$，$\mu\in R^1$，$\sigma^2>0$ 皆未知，x_1，x_2，…，$x_n(n\geqslant 2)$ 为从总体 X 的简单抽样，估计 $\sigma^a(a>0)$ 的损失函数为 $(d\sigma^{-a}-1)^2$，则

$$\hat{\sigma}^a=\begin{cases} c_0s^a, & \text{当}\quad (s^2Q^{-1})^{\frac{a}{2}}<[2c_0\bar{c}_a(0)]^{-1}; \\ Q^{\frac{a}{2}}g(s^aQ^{-\frac{a}{2}}), & \text{其它}. \end{cases} \tag{4.7}$$

一致优于最优同变估计 c_0s^a，因而 $\hat{\sigma}^a$ 也是 σ^a 的 minimax 估计，此处 $g(t)$ 由(4.6)式给出。

系 特取 $g(t)=[2c_0\bar{c}_a(0)]^{-1}$，可使(4.4)式中 $J=0$ 时条件期望达到最大。也就是 Stein 所得的结果

$$\hat{\sigma}^a=\min(c_0s^a,\ Q^{\frac{a}{2}}(\bar{c}_a(0))^{-1})$$

在 $a=1$ 时，得

$$\hat{\sigma}=\min\left(\frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{2}\,\Gamma\left(\frac{n+1}{2}\right)}s,\ \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{2}\,\Gamma\left(\frac{n+2}{2}\right)}Q^{\frac{1}{2}}\right);$$

在 $a=2$ 时，得

$$\hat{\sigma}^2=\min((n+1)^{-1}s^2,\ (n+2)^{-1}Q).$$

还应指出(4.7)式所给的估计并非一致优于 c_0s^2 的全部估计量. 例如，$a=2$ 时

$$\hat{\sigma}^2=\int_0^{n\bar{x}^2s^{-2}} s^2(1+t)^{-\frac{n+2}{2}}t^{-\frac{1}{2}}dt\Big/$$

$$\int_0^{n\bar{x}^2s^{-2}} (n+2)(1+t)^{-\frac{n+4}{2}}t^{-\frac{1}{2}}dt \qquad (4.7)^*$$

也一致优于$(n+1)^{-1}s^2$，它在刻度变换的同变估计类中是容许的[6]，但不属于(4.7)所定义的估计类.

用同样思路可以把此结果推广到 k 维正态分布 $N(\mu,\ A)$ 的广义方差 $\det A$ 的估计，相应于 Stein 估计为

$$\det\hat{A}=\min\left\{\frac{(n-k+1)!}{(n+1)!}\det s^2,\ \frac{(n-k+2)!}{(n+2)!}\det(s^2+n\bar{x}\bar{x}')\right\}, \qquad (4.8)$$

此处
$$s^2=\sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x})'.$$

由于 $k>1$ 的协方差估计，第七章 § 2 求出的最优同变估计就不同于常用的 c_ns^2，而后者也是仿射群下的同变估计，从而常用的皆不是容许的.

二、三维以上均值估计的改进

若 $X=(X_1,\ \cdots,\ X_p)'$ 为随机向量，$X\sim N(\mu,\ I_p)$，不难计算 $E\|X\|^2=\|\mu\|^2+p$. 由此可见，用 X 估计 μ 是大了一点，特别是 p 比较大一些，就更明显. 如能进行压缩，就有可能找到使均方风险更小的估计. 如何压缩呢？可以设想 μ 有先验分布 $N(0,bI_p)$，不难在$\|\mu-d\|^2$ 平方损失下求得 μ 的 Bayes 估计 $\hat{\mu}=bx(1+b)^{-1}$，

但 b 往往不知道，注意 $E\|\mu\|^2=bp$ 以及 $E\|X\|^2=\|\mu\|^2+p$，可用 $\|\mu\|^2p^{-1}$ 来估计 b，又用 $\|X\|^2-p$ 来估计 $\|\mu\|^2$，即 b 的估计 $\hat{b}=(\|x\|^2-p)p^{-1}$，从而求得 μ 的经验 Bayes 估计为

$$\hat{\mu}^*=\frac{\hat{b}}{1+\hat{b}}x=\left(1-\frac{p}{\|x\|^2}\right)x.$$

这就给出了对 x 的一个压缩. James 和 Stein 找到了如此形式的估计 $\hat{\mu}(x)=(1-(p-2)\|x\|^{-2})x$，并证明了一致优于常用估计 x，近 20 年来有许多作者进一步研究了哪一些估计一致优于常用估计 x，得到各种形式的结果. 一种更普遍使用的形式是把 $\hat{\mu}(x)$ 中 $p-2$ 的因子换为 $cr(\|x\|^2)$ 的形式，而且从这种形式中，在 $p\geqslant5$ 时，找到了一些 μ 的 Bayes 容许 minimax 估计. 在 $p=3$、4 时，如此形式的 Bayes 容许 minimax 不存在[4]，但容许 minimax 估计存在[1,15]. 下面我们采用恒等式的方法来证明上述有关结果[17].

引理 4.1 设 r. v $X\sim N(\mu, A)$，A 为 $p\times p$ 正定阵. 若 $h(x)$ 的一阶偏导数存在，$E\left(\left\|\frac{\partial h(x)}{\partial x}\right\|\Big|\mu, A\right)<\infty$，

$$\lim_{\|x\|\to\infty}|h(x)|e^{-\frac{1}{2}(x-\mu)A^{-1}(x-\mu)}=0,$$

则

$$E_\mu((X-\mu)'h(x))=E_\mu\left(t_r A\frac{\partial}{\partial x}h(x)\right). \tag{4.9}$$

证

$$E_\mu((X-\mu)'h(X))=(2\pi)^{-\frac{p}{2}}(\det A)^{-\frac{1}{2}}$$
$$\times\int_{R^p}(x-\mu)'h(x)e^{-\frac{1}{2}(x-\mu)'A^{-1}(x-\mu)}dx.$$

因为存在 B 使得 $B'B=A$，令 $y-\tilde{\mu}=(B')^{-1}(x-\mu)$，$\tilde{\mu}=(B')^{-1}\mu$，就得

$$E_\mu((X-\mu)'h(X))$$
$$=(2\pi)^{-\frac{p}{2}}\int_{R^p}(y-\tilde{\mu})'Bh(B'y)e^{-\frac{1}{2}\|y-\tilde{\mu}\|^2}dy$$

$$= (2\pi)^{-\frac{p}{2}} \sum_{i,j}^{p} \int_{R^p} (y_i - \tilde{\mu}_i) b_{ij} h_j(B'y) e^{-\frac{1}{2}\|y-\tilde{\mu}\|^2} dy$$

$$= (2\pi)^{-\frac{p}{2}} \sum_{i,j=1}^{p} \int_{R^p} b_{ij} \frac{\partial}{\partial y_i} h_j(B'Y) e^{-\frac{1}{2}\|y-\tilde{\mu}\|^2} dy$$

$$= (2\pi)^{-\frac{p}{2}} (\det A)^{-\frac{1}{2}}$$

$$\times \int_{R^p} \sum_{i,j=1}^{p} \sum_{k=1}^{p} b_{ij} b_{ik} \frac{\partial}{\partial x_k} h_j(x) e^{-\frac{1}{2}(x-\mu)'A^{-1}(x-\mu)} dx$$

$$= E_\mu\left(\sum_{k,j=1}^{p} a_{kj} \frac{\partial}{\partial x_k} h_j(x) \right) = E_\mu\left(t_r A \frac{\partial}{\partial x} h(x) \right). \quad (4.10)$$

(4.10) 中 $(h_1(x), \cdots, h_p(x))' = h(x)$, $x = (x_1, \cdots, x_k)$.

$$\frac{\partial}{\partial x} h(x) = \left(\frac{\partial h_j(x)}{\partial x_k} \right)_{p\times p}, \quad A = (a_{ij})_{p\times p}, \; B = (b_{ij})_{p\times p}.$$

(4.10) 中的第三个等式是利用 y_i 的分部积分法得到的，第四个等式是利用 $B'B = A$ 得到的. 至此引理证毕.

公式(4.9)是多维正态下的恒等式. 此恒等式是目前解决参数估计容许性问题的有力工具之一. 读者从习题中还可再看到它的作用.

定理 4.2 设 r. v $X \sim N(\mu, A)$, $A > 0$, 若 $r(t)$ 在 $(0, \infty)$ 可微且 $r'(t) \geqslant 0$, $0 \leqslant r(t) \leqslant (p-2)2$ (当 $t \in (0, \infty)$), 则

$$\delta(x) = (I_p - r(x'A^{-2}x)(x'A^{-2}x)^{-1}A^{-1})x \quad (4.11)$$

是 μ 的 minimax 估计，若 $r(t)$ 在 $(0, \infty)$ 上不恒为零或 $2(p-2)$, 则 $\delta(x) \prec x$. $\delta(x)$ 与 x 风险之差为

$$E_\mu(\|X-\mu\|^2) - E_\mu(\|\delta(X) - \mu\|^2)$$
$$= E_\mu\{ r(X'A^{-2}X)(X'A^{-2}X)^{-1}[2(p-2) - r(X'A^{-2}X)]$$
$$+ 4r'(X'A^{-2}X)\} \geqslant 0 \quad (\text{当 } \mu \in R^p). \quad (4.12)$$

证 为方便起见，我们记 $Y = X'A^{-2}X$

$$E_\mu(\|X-\mu\|^2 - \|X - \mu - r(Y)Y^{-1}A^{-1}X\|^2)$$
$$= -E_\mu(r^2(Y)Y^{-1}) + 2E_\mu[(X-\mu)'A^{-1}Xr(Y)Y^{-1}]. \quad (4.13)$$

取 $h(x) = r(y)y^{-1}A^{-1}x = r(x'A^{-2}x)(x'A^{-2}x)^{-1}A^{-1}x$, 那么

$$t_r\left[A\left(\frac{\partial h}{\partial x}\right)\right]$$
$$=py^{-1}r(y)+2x'A^{-2}x((x'A^{-2}x)^{-1}r'(y)-r(y)y^{-2})$$
$$=y^{-1}r(y)(p-2)+2r'(y). \tag{4.14}$$

根据引理 4.1, $E_\mu(X-\mu)'h(X)=E_\mu t_r\left(A\left(\frac{\partial h}{\partial x}\right)\right)$, 将(4.14)的结果代入此式，然后再代入(4.13)式，就得(4.12)等式. 由于 $0\leqslant r(t)\leqslant 2(p-2)$, $r'(t)\geqslant 0$, $t\in(0,\infty)$, 故(4.12)的不等号显然成立. 而在第七章§3中已证明了 x 是 μ 的 minimax 估计. 由(4.12)式即知 $\delta(x)$ 也是 μ 的 minimax 估计. 当 $r(t)$ 不恒等于零和 $2(p-2)$ 时, 显然(4.12)严格不等号成立. 因此 $\delta(X)\prec X$.

下面将在 $p\geqslant 5$ 时，给出 μ 的 Bayes 容许 minimax 估计. 在 $p=3$、4 时给出 μ 的容许 minimax 估计.

我们记 $a_1, \cdots, a_p$ 为 A 的特征根, $b_i(\lambda)=a_i(a_i-\lambda)\lambda^{-1}(i=1, 2, \cdots, p)$, $\alpha=\min\limits_{1\leqslant i\leqslant p} a_i$, $c<2$, $a\geqslant 0$, 选用如下先验测度，它具有关于 Lebesgue 测度的密度

$$g_{b,c}(\mu)=\int_0^\alpha\left(\prod_1^p b_i(\lambda)\right)^{-1/2}$$
$$\times\exp\left\{-\frac{1}{2}\mu'A^{-2}(I-\lambda A^{-1})^{-1}\mu\lambda\right\}(\lambda+b)^{-c}d\lambda. \tag{4.15}$$

关于 $g_{a,c}(\mu)$ 的 Bayes 估计或 $GBDF$ 为

$$\delta_{b,c}(x)=\int_{R^p}\mu g_{a,c}(\mu)e^{-(x-\mu)'A^{-1}(x-\mu)/2}d\mu$$
$$\times\left(\int_{R^p}g_c(\mu)e^{-(x-\mu)'A^{-1}(x-\mu)}d\mu\right)^{-1}. \tag{4.16}$$

经过配方正态积分计算，可得

$$\int_{R^p}\mu\exp\left\{-\frac{\lambda}{2}\mu'A^{-2}(I-\lambda A^{-1})^{-1}\mu\right.$$
$$\left.-\frac{1}{2}(\mu-x)'A^{-1}(\mu-x)\right\}d\mu$$
$$=\int_{R^p}\mu\exp\left\{-\frac{1}{2}(\mu-(I-\lambda A^{-1})x)'(\lambda A^{-2}(I-\lambda A^{-1})^{-1}\right.$$

$$+A^{-1})(\mu-(I-\lambda A^{-1})x)-\frac{\lambda}{2}x'A^{-2}x\Big\}d\mu$$

$$=(2\pi)^{p/2}\Big(\prod_1^p(a_i-\lambda)^{1/2}a_i^{1/2}\Big)e^{-\lambda x'A^{-2}x/2}(I-\lambda A^{-1})x, \quad (4.17)$$

$$\int_{R^p}\exp\Big\{-\frac{\lambda}{2}\mu'A^{-2}(I-\lambda A^{-1})^{-1}\mu-\frac{1}{2}(\mu-x)'A^{-1}(\mu-x)\Big\}d\mu$$

$$=(2\pi)^{p/2}\Big(\prod_1^p(a_i-\lambda)^{1/2}a_i^{1/2}\Big)e^{-\lambda x'A^{-2}x/2}. \quad (4.18)$$

将(4.17)、(4.18)代入(4.16)式右端，即得广义 Bayes 解

$$\delta_{b,c}(x)=\Big\{I-\int_0^\alpha\lambda^{p/2+1}(\lambda+b)^{-c}e^{-\frac{\lambda}{2}x'A^{-2}x}d\lambda$$

$$\times\Big(\int_0^\alpha\lambda^{p/2}(\lambda+b)^{-c}e^{-\lambda x'A^{-2}x/2}d\lambda\Big)^{-1}A^{-1}\Big\}x, \quad (4.19)$$

故

$$r_{b,c}(t)=tI_{b,c}\Big(\frac{p}{2}+1,\ t\Big)I_{b,c}^{-1}\Big(\frac{p}{2},\ t\Big)$$

$$=p+2-2\Big(\alpha^{p/2+1}(\alpha+b)^{-c}e^{-\alpha t/2}$$

$$+c\int_0^\alpha\lambda^{p/2+1}(\lambda+b)^{-c-1}e^{-\lambda t/2}\Big)I_{b,c}^{-1}\Big(\frac{p}{2},\ t\Big), \quad (4.20)$$

此处

$$I_{b,c}(p/2,\ t)=\int_0^\alpha\lambda^{p/2}(\lambda+b)^{-c}e^{-\lambda t/2}d\lambda. \quad (4.21)$$

当 $b=0,\ c<1$ 时，

$$\int_{R^p}g_{b,c}(\mu)d\mu=(2\pi)^{p/2}(1-c)^{-1}\alpha^{-c+1}<\infty.$$

因此 $\delta_{0,c}(x)$ 是 Bayes 估计，又因 $r'_{0,c}(t)>0,\ 0<r_{0,c}(t)\leqslant p+2-2c$. 如果 $p+2-2c\leqslant 2(p-2)$ (即 $c\geqslant 3-p/2,\ p\geqslant 5$) 由定理 4.2 及平方损失下 Bayes 估计的唯一性，$\delta_{0c}(x)$ 是 Bayes 容许 minimax 估计. 而在 $b=0,\ 3-p/2\leqslant c<2\ (p=3,\ 4)$ 时 $\delta_{0,c}(x)$ 是广义 Bayes 估计. 根据定理 4.2，它是优于 x. 因此是 minimax 估计. 下面证明它的容许性.

$$\frac{d}{dt}r_{b,c}(t)=t^{-1}r_{b,c}(t)+2^{-1}t^{-1}r_{b,c}^2(t)$$

$$-2^{-1}tI_{b,c}(2+p/2,\ t)I_{b,c}^{-1}(p/2,\ t). \quad (4.22)$$

将此式代入(4.12)，并记$Y=X'A^{-2}X$，就有

$$\begin{aligned}D_\mu(\delta_{b,c})&\triangleq E_\mu\{\|X-\mu\|^2-\|\delta_{b,c}(X)-\mu\|^2\}\\&=E_\mu\{r_{b,c}(y)y^{-1}(2p+r_{b,c}(y))\\&\quad-2yI_{b,c}(2+p/2,\ y)I_{b,c}^{-1}(p/2,\ y)\}.\end{aligned}\tag{4.23}$$

再利用(4.18)式及积分次序交换，就可得到

$$\begin{aligned}\widetilde{E}_{b,c}(D_\mu(\delta_{b,c}))&\triangleq\int_{R^p}D_\mu(\delta_{b,c})g_{b,c}(\mu)d\mu\\&=\prod_1^p a_i^{-1/2}\int_{R^p}[I_{b,c}(1+p/2,\ Y)(2p+r_{b,c}(y))\\&\quad-2YI_{b,c}(2+p/2,\ Y)]dx\\&=2^{p/2}\Gamma(p/2)\prod_1^p a_i^{1/2}\int_0^\infty[I_{b,c}(1+p/2,\ z)(2p+r_{b,c}(z))\\&\quad-2zI_{b,c}(2+p/2,\ z)]z^{p/2-1}dz\\&=2^{p/2}\Gamma(p/2)\prod_1^p a_i^{1/2}\int_0^\infty I_{b,c}(1+p/2,\ z)r_{b,c}(z)dz.\end{aligned}\tag{4.24}$$

上式第三个等式是先作变换$A^{-1}x=\xi$，再把ξ化为极坐标计算积分而得，最后的等式是利用积分交换及恒等式

$$\int_0^\infty z^{p/2}e^{-z/2}dz=p\int_0^\infty z^{p/2-1}e^{-z/2}dz$$

而得.

由(4.20)知$0<r_{b,c}(z)\leqslant p+2$、$I_{b,c}(1+p/2,\ z)$是$b$的单调下降函数，利用单调及控制收敛定理，有

$$\lim_{b\to0^+}\widetilde{E}_{b,c}(D_\mu(\delta_{b,c}))=\widetilde{E}_{0,c}(D_\mu(\delta_{0,c}))<\infty.\tag{4.25}$$

因为$b>0$，$\int_{R^p}g_{b,c}(\mu)<\infty$，$\delta_{b,c}(x)$是$\mu$的 Bayes 估计，也是 OGBDF. 由定理 1.4 及 (4.25)，$\Delta_b(\delta_{0c},\ \delta_{b,c})=\widetilde{E}_{b,c}(D_\mu(\delta_{0,c})-D_\mu(\delta_{b,c}))\to0$（当$b\downarrow0$），故$\delta_{0,c}(x)$是容许估计.

综上所述，便得如下定理：

定理 4.3 若$X\sim N(\mu,\ A)$，$\mu\in R^p$未知，$A>0$已知，则

$$\delta_{0,c}=\left(I-\int_0^\alpha\lambda^{p/2+1-c}e^{-\lambda x'A^{-1}x/2}d\lambda\left(\int_0^\alpha\lambda^{p/2-c}e^{-\lambda x'A^{-1}x}d\lambda\right)^{-1}A^{-1}\right)x\tag{4.26}$$

在平方损失下，当 $3-p/2\leqslant c<1$ 时，它是 μ 的 Bayes 容许 minimax 估计，而在 $3-p/2\leqslant c<2$ 时，它是 μ 的容许 minimax 估计.

现在来讨论一下，平方损失与一般二次型损失下对容许估计的影响. 若 $X\sim P_\theta$, $\theta\in\Theta$ 估计参数 θ 的损失为 $(d-\theta)'B(d-\theta)$ $(B\geqslant 0)$，若 $d(x)$ 是 θ 在上述损失的容许估计，为了强调与 B 的关系，记为 $d(x)\overset{\text{ad. }B}{\sim}\theta$.

定理 4.4 若 $B_0>0$, $t(x)\overset{\text{ad. }B_0}{\sim}\theta$, 则对任意的 $B\geqslant 0$，有 $t(x)\overset{\text{ad. }B}{\sim}\theta$.

证 不失一般性，令 $B_0=I$，用反证法：若对某个 $B\geqslant 0$ $(B\neq 0)$，$t(x)\overset{\text{ad. }B}{\sim}\theta$ 不成立，则存在 $t_1(x)$，有

$$E_\theta(t_1-\theta)'B(t_1-\theta)\leqslant E(t-\theta)'B(t-\theta) \tag{4.27}$$

对一切 $\theta\in\Theta$ 成立，且对某些 θ 严格不等号成立. 以 λ 记 B 的最大特征根，显然 $\lambda>0$ $(\because\ B\neq 0)$，记 $F=B/\lambda$，显然 $0\leqslant F\leqslant I$，因此 $F^2\leqslant I$，作估计 $\tilde{t}(x)=t(x)+F(t_1(x)-t(x))$，则

$$\begin{aligned}
&E(\tilde{t}-\theta)'(\tilde{t}-\theta)\\
&=E(t-\theta)'(t-\theta)+E(t_1-t)'F^2(t_1-t)\\
&\quad+E(t-\theta)'F(t_1-t)+E(t_1-t)'F(t-\theta)\\
&\leqslant E(t-\theta)'(t-\theta)+E(t_1-t)'F(t_1-t)\\
&\quad+E(t-\theta)F(t_1-t)+E(t_1-t)'F(t-\theta)\\
&=E(t-\theta)'(t-\theta)+E(t_1-\theta)'F(t_1-\theta)\\
&\quad-E(t-\theta)'F(t-\theta)\\
&\leqslant E(t-\theta)'(t-\theta)\quad(\text{当 }\theta\in\Theta).
\end{aligned} \tag{4.28}$$

而且最后一个不等式对某些 θ 严格不等式成立，故与 $t\overset{\text{ad. }I}{\sim}\theta$ 相矛盾. 因此定理得证.

由此定理保证了定理 4.3 对二次型损失也成立. 因此，以后在二次型损失函数中，对于容许性问题，只考虑平方损失就够了.

三、线性回归模型

下面把定理 4.2 作一点推广，使之能运用到线性回归模型上，从而改进最小二乘估计.

引理 4.2 若 $T(w)$ 为 $(0, \infty)$ 上的单调非增非负函数，w 为非负随机变量，则

$$ET(w)(Ew-w)\geqslant 0.$$

证 记 w 的分布函数为 $P(w)$，则

$$ET(w)(Ew-w)=\left(\int_0^{Ew}+\int_{Ew}^{\infty}\right)T(w)(Ew-w)dP(w)$$

$$\geqslant T(Ew)\int_0^{\infty}(Ew-w)dP(w)=0.$$

定理 4.4 设 $X\sim N(\mu, \sigma^2A)$，$A>0$ 已知，$\mu\in R^p$，$\sigma>0$ 皆未知，损失函数为 $\|\mu-d\|^2/\sigma^2$，若 $0\leqslant r(t)\leqslant 2(p-2)/(n+2)$，$r'(t)\geqslant 0$，当 $t\in(0, \infty)$，S_n^2 与 X 独立，且 $S_n^2/\sigma^2\sim x_n^2$ 分布，则

$$\hat{\mu}(x, S_n^2)=[I-r(x'A^{-2}xS_n^{-2})(x'A^{-2}x)^{-1}S_n^2A^{-1}]x \quad (4.29)$$

是 μ 的 minimax 估计. 当 $r(t)$ 不恒为 0 和 $2(p-2)/(n+2)$ 时，则 $\hat{\mu}(x, S_n^2)\prec X$，改进了 μ 的估计 X.

证 由于 x 与 $\hat{\mu}(x, S_n^2)$ 皆是刻度变换下的同变估计，故

$$\sigma^{-2}E_{\sigma,\mu}\{\|x-\mu\|^2-\|\hat{\mu}(x, S_n^2)-\mu\|^2\}$$

$$=E_{1,\tilde{\mu}}\{\|X-\tilde{\mu}\|^2-\|\hat{\mu}(X, S_n^2)-\mu\|^2\}.$$

此处 $\tilde{\mu}=\mu/\sigma$，故不妨设 $\sigma=1$. 由于 S_n^2 与 X 独立及恒等式 (4.11)（请注意，此时每个 $r(\cdot)$ 前乘上 S_n^2，而 $r'(\cdot)$ 则不乘 S_n^2）以及 χ^2 分布恒等式 $Eh(\chi_n^2)=2nE(h(\chi_{n+2}^2)/\chi_{n+2}^2)$（此处 χ_n^2 表示自由度为 n 的 χ^2 变量），便有

$$\begin{aligned}
&E_{\mu}\{\|X-\mu\|^2-\|\hat{\mu}(X, S_n^2)-\mu\|^2\}\\
&=E_{\mu}\{r(X'A^{-2}XS_n^{-2})(X'A^{-2}X)^{-1}S_n^2[2(p-2)\\
&\quad -r(X'A^{-2}XS_n^{-2})S_n^2]+4r'(X'A^{-2}XS_n^{-2})\}\\
&=E_{\mu}\{2nr(X'A^{-2}XS_{n+2}^{-2})(X'A^{-2}X)^{-1}(2(p-2)\\
&\quad -r(X'A^{-2}XS_{n+2}^{-2})S_{n+2}^2)\}+E_{\mu}\{4r'(X'A^{-2}XS_n^{-2})\}
\end{aligned}$$

$$\geqslant E_{\mu}\{2nr(X'A^{-2}XS_{n+2}^{-2})(X'A^{-2}X)^{-1}(2(p-2)$$
$$-r(X'A^{-2}XS_{n+2}^{-2})(n+2)\}+4E_{\mu}\{r'(X'A^{-2}XS_{n}^{-2})\}$$
$$\geqslant 0 \quad (当 \mu\in R). \tag{4.30}$$

上式第一个不等号是利用引理 4.2, 其中 $S_{n+2}^2\sim\chi_{n+2}^2$ 分布与 X 独立. (4.30)最后的不等号是根据假设,而且在 $r(t)$ 在$(0, \infty)$上不恒为零和 $\dfrac{2(p-2)}{n+2}$, 则严格不等号成立. 说明 $\hat{\mu}(X, S_n^2)\prec X$. 在第七章 § 4 中已证明 X 是 minimax 估计, 因此 $\hat{\mu}(X, S_n^2)$ 也是.

若 $Y_n=X_n\beta+\varepsilon_n$, $\varepsilon_n\sim N(0, \sigma^2 I_n)$, $\beta\in R^p$, $\sigma>0$ 皆未知 ($n>p$), 且 $X_n'X>0$. 这是熟知的回归模型, β 的最小二乘估计 $\hat{\beta}_L=(X_n'X_n)^{-1}X_n'Y_n\sim N(\beta, \sigma^2(X_n'X_n)^{-1})$, 其残差平方和 $S^2=Y_n'(I-X_n(X_n'X_n)^{-1}X_n')Y_n$, $S^2/\sigma^2\sim\chi_{n-p}^2$ 分布; 利用定理 4.4, 即可推得如下定理:

定理 4.5 在上述回归模型中, 若 $r(t)$ 满足 $0\leqslant r(t)\leqslant 2(p-2)/(n-p+2)$, $r'(t)\geqslant 0$, 且不恒为零和 $2(p-2)/(n-p+2)$, 则

$$\hat{\beta}=\{I-r(\hat{\beta}_L'(X_n'X_n)^2\hat{\beta}_L S^{-2})(\hat{\beta}_L'(X_n'X_n)^2\hat{\beta}_L)^{-1}S^2X_n'X_n\}\hat{\beta}_L$$

在 $L(\beta, d)=\sigma^{-2}\|\beta-d\|^2$ 下, $\hat{\beta}\prec\hat{\beta}_L$, 且 $\hat{\beta}$ 是 β 的 minimax 估计.

σ^2 的最优同变估计 $S^2/(n-p+2)$ 也是非容许的.

$$\hat{\sigma}_n^2=\min\left\{(n-p+2)^{-1}S^2, \sum_{i=1}^n y_i^2(n+2)^{-1}\right\}$$

就一致优于它.

下面我们应指出, 如果定理 4.4 中没有有关 S_n^2 的假设或者重复抽样假设, 那么情况完全两样.

定理 4.6 若 p 维随机向量 $X\sim N(\mu, \sigma^2 A)$, $A>0$ 已知; $\mu\in R^p$, $\sigma>0$ 未知, 则 X 在平方损失下是 μ 的容许估计.

证 若存在一估计 $\delta(x)\prec x$, 则

$$E_{\theta,\sigma}\|\delta(X)-\theta\|^2\leqslant E_{\theta,\sigma}\|X-\theta\|^2=t_r A\sigma^2$$

当 $\theta\in R^p$, $\sigma>0$. 而且至少有一个点(θ_0, σ_0), 使得上式严格不等号

成立．令 $B=\{x;\ \delta(x)\neq x\}$，显然 B 的 Lebesgue 测度 $L(B)>0$．因此存在 $p_0>0$，使得

$$L(x:\ \|\delta(x)-x\|^2>p_0^{-1})\triangleq L(B_{p_0})>0,\quad B_{p_0}\subseteq B.$$

设 M 是 B_p 中任意的 Lebesgue 点．则

$$E_{M,\sigma}(\|\delta(X)-X\|^2)$$

$$\geqslant p_0^{-1}(\det A)^{-1/2}\int_{B_{p_0}}\varphi\left(\frac{(x-M)'A^{-1}(x-M)}{\sigma}\right)\sigma^{-p}dx,$$

此处 $\varphi(t)=\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}$．因此存在 $\delta_1>0$，$\delta_2>0$，使得

$$\varphi\{(x-M)'A^{-1}(x-M)/\sigma\}>\delta_1,$$

当
$$|x-M|=\max_{1<i<p}|x_i-M_i|<\delta_2\sigma.$$

根据 Lebesgue 点的定义，则有

$$\liminf_{\sigma\to0}E_{M,\sigma}(\|\delta(X)-X\|^2)$$

$$\geqslant\liminf_{\sigma\to0}p_0^{-1}\delta_1(\det A)^{-\frac{1}{2}}\int_{|x-M|<\delta_2\delta}I_{B_{p_0}}(x)\sigma^{-k}dx$$

$$=p_0^{-1}\delta_1(2\delta_2)^{p}\cdot(\det A)^{-\frac{1}{2}}>0.\tag{4.31}$$

另一方面

$$E_{M,\sigma}(\|\delta(X)-X\|^2)$$

$$\leqslant 2E_{M,\sigma}(\|\delta(X)-M\|^2)+2E_{M,\sigma}\|X-M\|^2$$

$$\leqslant 4t_rA\sigma^2\to0\quad(当\ \sigma\to0).$$

此式与(4.13)式发生了矛盾．反证假设不成立．定理得证．

四、多维正态分布均值的线性容许估计类

若 $X\sim N(\mu,\ A)$，$A>0$ 已知；$\mu\in R^p$ 未知，现在问哪些线性估计 $C_{p\times p}X$，$C_{p\times p}X+d$ 是 μ 在平方损失 $\|\mu-d\|^2$ 下是容许的？哪些不是容许的？下面的引理和定理在不加说明的情况下，上述假设皆成立．

引理 4.3 若 P 为正交阵，$A=I_p$，则

$$CX \overset{\text{ad. } I}{\sim} \mu \Leftrightarrow P'CPX \overset{\text{ad. } I}{\sim} \mu.$$

证 因为 $PX \sim N(P\mu, I_p)$，所以

$$CX \overset{\text{ad. } I}{\sim} \mu \Leftrightarrow CPX \overset{\text{ad. } I}{\sim} P\mu.$$

又因为 $\|P\mu - PCX\|^2 = \|\mu - CX\|^2$，所以

$$CPX \overset{\text{ad. } I}{\sim} P\mu \Leftrightarrow P'CPX \overset{\text{ad. } I}{\sim} P'P\mu = \mu.$$

上述两个充要条件便证得引理.

定理 4.7(cohen [11]) 若 $X \sim N(\mu, I_p)$，则 $CX \overset{\text{ad. } I}{\sim} \mu$，$C$ 对称，其特征根在[0, 1]上，而且等于1的特征根最多为2个.

证 首先证明若 C 是非对称的，则 CX 为 μ 的非容许估计. 作

$$C^* = I_p - [(C - I_p)'(C - I_p)]^{\frac{1}{2}}, \tag{4.82}$$

必有 $C^*X \prec CX$. 我们注意到

$$(C^* - I_p)'(C^* - I_p) = (C - I_p)'(C - I_p),$$

$$E_\mu\|CX - \mu\|^2 = \text{tr}\, C'C + \mu'(C - I_p)'(C - I_p)\mu,$$

$$\begin{aligned} E_\mu\|C^*X - \mu\|^2 &= \text{tr}(C^*)'C^* + \mu'(C^* - I_p)'(C^* - I_p)\mu \\ &= \text{tr}(C^*)'C^* + \mu'(C - I_p)'(C - I_p)\mu, \end{aligned}$$

$$(C^*)'C^* = 2I_p - C - C' - 2[(C - I_p)'(C - I_p)]^{\frac{1}{2}} + C'C.$$

故 $\text{tr}(C^*)'C^* < \text{tr}\, C'C \Leftrightarrow \text{tr}(I - C) < \text{tr}[(I_p - C)'(I_p - C)]^{\frac{1}{2}}.$

令 $B = I_p - C$，则由矩阵论中的正交特征分解定理知，存在正交阵 P、Q，使

$$B = P'\begin{pmatrix} \Lambda & 0 \\ 0 & 0 \end{pmatrix}Q,\ \Lambda = \text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_q)\ (q \leqslant p).$$

记 $\bar{P} = QP'$ 仍为正交阵，其对角线上的元素记为 $\bar{p}_{ii}$，且 $|\bar{p}_{ii}| \leqslant 1$ $(i = 1, 2, \cdots, p)$，不难计算得到

$$\text{tr}B = t_r\begin{pmatrix} \Lambda & 0 \\ 0 & 0 \end{pmatrix}QP' = \sum_{i=1}^{q} \lambda_i \bar{p}_{ii} \leqslant \sum_{i=1}^{q} |\lambda_i| = \text{tr}(B'B)^{\frac{1}{2}}.$$

上式等号成立的充要条件是 $\bar{p}_{ii} = 1 (i = 1, 2, \cdots, q)$. 由于 QP' 为

正交阵，故 $\bar{P}=\begin{pmatrix} I_q & 0 \\ 0 & * \end{pmatrix}$. 此时

$$B=P'\begin{pmatrix} \Lambda & 0 \\ 0 & 0 \end{pmatrix}Q=Q'\bar{P}\begin{pmatrix} \Lambda & 0 \\ 0 & 0 \end{pmatrix}Q$$

$$=Q'\begin{pmatrix} I_q & 0 \\ 0 & * \end{pmatrix}\begin{pmatrix} \Lambda & 0 \\ 0 & 0 \end{pmatrix}Q=Q'\begin{pmatrix} \Lambda & 0 \\ 0 & 0 \end{pmatrix}Q.$$

由此知道 B 是对称阵. 这与 C 为非对称相矛盾. 故

$$\operatorname{tr}B<\operatorname{tr}(B'B)^{\frac{1}{2}},$$

从而 $\operatorname{tr}(C^*)'C^*<\operatorname{tr}C'C$, 以致 $E_\mu\|C^*X-\mu\|^2<E_\mu\|(X-\mu)\|^2$. 故 CX 是非容许的.

所以下面总设 C 为对称阵. 故存在正交阵 P, 使 $P'CP=$ 对角阵. 由引理 4.3 知，讨论 CX 的容许性，不妨仅考虑 C 为对角阵的情况，其中对角线上的元素记为 $c_1, c_2, \cdots, c_p$. 容易算得

$$R(\mu, CX)=E\|\mu-CX\|^2=\sum_{i=0}^{p}(c_i^2+\mu_i^2(1-c_i)^2). \quad (4.33)$$

必要性：由(4.33)式可见，若有 $c_i<0$, 则把 c_i 换成零；若有 $c_i>1$, 则把 c_i 换为 1. 这样得到的 μ 的估计 $\tilde{C}X$ 显然一致优于原来的 CX, CX 就是非容许估计. 因此，欲使 CX 为容许，必有 $0\leqslant c_i\leqslant 1 (i=1, 2, \cdots, p)$. 另一方面，若 $c_1, \cdots, c_p$ 中有三个元素为 1. 不妨设 $c_1=c_2=c_3=1$. 造一个估计

$$\hat{\mu}=\left(\left(1-\frac{1}{\sum_1^3 x_i^2}\right)(x_1, x_2, x_3), c_4x_4, \cdots, c_px_p\right)'. \quad (4.34)$$

由(4.11)式知

$$E\|\hat{\mu}-\mu\|^2=3-E(X_1^2+X_2^2+X_3^2)^{-1}$$

$$+\sum_{i=4}^{p}(\mu_i^2(1-c_i)^2+c_i^2)<E\|CX-\mu\|^2.$$

故 CX 为非容许. 因此，$c_1, c_2, \cdots, c_p$ 中只能有两个为 1, 方能使 CX 为 μ 的容许估计. 必要性得证.

充分性：设 $C=\operatorname{diag}(c_1, \cdots, c_p)$, 若 $c_i=0$, 定义 μ_i 的先验分布为 $P(\mu_i=0)=1$; 若 $c_i>0$, 则定义 μ_i 的先验分布为

$N\left(0, \frac{c_i}{1-c_i}\right)$，且诸 μ_i 独立．则容易得出 μ 的 Bayes 估计为 $E(\mu|x)=CX$．由于平方损失(严凸)Bayes 解唯一，因此知 CX 是 μ 的容许估计．

若诸 c_i 中仅有一到两个 $c_i=1$，不妨设 $c_1=1$ 或 $c_1=c_2=1$．现在考虑后一种情形：此时若 CX 不是 μ 的容许估计．则存在估计 $\varphi(x)\prec CX$．记

$$\mu^{(1)}=\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix},\quad \mu=\begin{pmatrix}\mu^{(1)}\\ \mu^{(2)}\end{pmatrix},\quad X^{(1)}=\begin{pmatrix}x_1\\ x_2\end{pmatrix},$$

$$X=\begin{pmatrix}x^{(1)}\\ x^{(2)}\end{pmatrix},\quad \varphi^{(1)}(x)=\begin{pmatrix}\varphi_1(x)\\ \varphi_2(x)\end{pmatrix},\quad \varphi(x)=\begin{pmatrix}\varphi^{(1)}(x)\\ \varphi^{(2)}(x)\end{pmatrix},$$

$$C^{(1)}=\mathrm{diag}(c_1, c_2),\quad C^{(2)}=\mathrm{diag}(c_3, \cdots, c_p).$$

如同前面，取 $\mu^{(2)}$ 的先验分布为：$P(\mu_i=0)=1$，当 $c_i=0$，$\mu_i\sim N\left(0, \frac{c_i}{1-c_i}\right)$，当 $c_i>0$，且诸 μ_i 独立，若再取 $P(\mu^{(1)}=\mu_0^{(1)})=1$，$\mu_0^{(1)}$ 为 R^2 中任意确定的点．则 μ 的 Bayes 估计为$\begin{pmatrix}\mu_0^{(1)}\\ c^{(2)}x^{(2)}\end{pmatrix}$或 $\mu^{(2)}$ 的 Bayes 估计为 $C^{(2)}X^{(2)}$．其 Bayes 风险为 $\sum\limits_{i=3}^{p}c_i$．因此，根据 Bayes 估计的定义，估计$\begin{pmatrix}\mu_0^{(1)}\\ \varphi^{(2)}(x)\end{pmatrix}$在上述先验分布的平均风险

$$EE_\mu\|\varphi^{(2)}(X)-\mu^{(2)}\|^2\geqslant\sum_{i=3}^{p}c_i. \tag{4.35}$$

另一方面，由于 $\varphi(x)\prec CX$，知

$$\begin{aligned}
&EE_\mu\|\varphi(X)-\mu\|^2\\
&\quad=EE_\mu\|\varphi^{(1)}(X)-\mu^{(1)}\|^2+EE_\mu\|\varphi^{(2)}(X)-\mu^{(2)}\|^2\\
&\quad\leqslant EE(\|CX-\mu\|^2)=E\left(2+\sum_{i=3}^{p}c_i^2+\mu_i^2(1-c_i)^2\right)\\
&\quad=2+\sum_{i=3}^{p}c_i.
\end{aligned} \tag{4.36}$$

从上式及(4.35)式知

$$EE_\mu\|\varphi^{(1)}(X)-\mu^{(1)}\|^2\leqslant 2. \tag{4.37}$$

注意此时 X 的绝对分布密度 $g(x, \mu_0^{(1)})$ 为

$$(2\pi)^{-1}\exp\left\{-\frac{1}{2}\|x^{(1)}-\mu_0^{(1)}\|^2\right\}E\left\{(2\pi)^{-\frac{p-3}{2}}\exp\left\{-\frac{1}{2}\|\mu^{(2)}-x^{(2)}\|^2\right\}\right.$$

$$\triangleq (2\pi)^{-1}\exp\left\{-\frac{1}{2}\|x^{(1)}-\mu_0^{(1)}\|^2\right\}f(x^{(2)}).$$

此处$f(x^{(2)})$与$\mu_0^{(1)}$无关，再由于$\mu_0^{(1)}$的任意性，可知$x^{(1)}$为关于$\mu_0^{(1)}$的完全充分统计量．加上损失是平方损失．故由$x^{(1)}$组成的估计构成完全类．由于$x^{(1)}$是$\mu_0^{(1)}$的容许估计*)，故(4.37)式对所有的$\mu_0^{(1)}\in R^2$成立，当且仅当等号成立．因而$\varphi^{(1)}(x)=x^{(1)}$ a. s. 再由(4.35)与(4.36)式知

$$EE_\mu\|\varphi^{(2)}(X)-\mu^{(2)}\|^2=\sum_{i=3}^{p}c_i=EE_\mu\|C^{(2)}X^{(2)}-\mu^{(2)}\|^2.$$

由于平方损失 Bayes 估计的唯一性，就有$\varphi^{(2)}(x)=c^{(2)}x^{(2)}$ a. s．因而$E_\mu\|\varphi(X)-\mu\|^2=E_\mu\|CX-\mu\|^2$．这就证明了$CX$是$\mu$的容许估计．充分性得证．整个定理证完．

现在我们把这定理作进一步推广：

定理 4.8 设 r. v $X\sim N(\mu, A)$，$A>0$ 已知，则 $CX+d\overset{\text{ad.}I}{\sim}\mu$ 的充要条件是：i) CA 对称；ii) C 的所有特征根在[0 1]上，且等于1的特征根至多两个；iii) d 属于由$C-I_p$的行向量所张成的线性子空间$\mathfrak{M}(C-I_p)$．

证 先证$d=0$的情况：令$y=A^{-1/2}X$，$\eta=A^{-\frac{1}{2}}\mu$．则$EY=\eta$，$VAR(Y)=I_p$．因为

$$\begin{aligned}E_\mu(\|CX-\mu\|^2|\mu, A)&=E(\|CA^{\frac{1}{2}}Y-A^{\frac{1}{2}}\eta\|^2|\eta, I_p)\\&=E(A^{-\frac{1}{2}}CA^{\frac{1}{2}}Y-\eta)'A(A^{-\frac{1}{2}}CA^{\frac{1}{2}}Y-\eta)|\eta, I_p).\end{aligned}$$

所以

$$CX\overset{\text{ad.}I}{\sim}\mu\Leftrightarrow A^{-\frac{1}{2}}CA^{\frac{1}{2}}Y\overset{\text{ad.}A}{\sim}\eta\Leftrightarrow A^{-\frac{1}{2}}CA^{\frac{1}{2}}Y\overset{\text{ad.}I}{\sim}\eta.$$

最后这个"⇔"是根据定理4.4．再根据定理4.7，$CX\overset{\text{ad.}I}{\sim}\mu$的充要条件是$A^{-\frac{1}{2}}CA^{\frac{1}{2}}$为对称，特征根在[0, 1]上，等于1的特征根

---

*) 参见本章§7定理7.2的推论或[18]．

至多有两个. $A^{-\frac{1}{2}}CA^{\frac{1}{2}}$ 对称等价于 CA 对称,而 $A^{-\frac{1}{2}}CA^{\frac{1}{2}}$ 的特征根与 C 的特征根相等. 此即 i)、ii)的条件. 在这种情况, 定理得证.

现在,条件 iii)对一般 $CX+d \overset{\text{ad. } I}{\sim} \mu$ 是必要的. 若 $d=(C-I_p)d_1+(C-I_p)^{\perp}d_2$, 那么

$$
\begin{aligned}
&E(\|CX+d-\mu\|^2|\mu,\ A)\\
&\quad=\operatorname{tr} CAC'+d_1'(C-I_p)'(C-I_p)d_1\\
&\qquad+2d_1'(C-I_p)'(C-I_p)\mu+\mu'(C-I_p)'(C-I_p)\mu\\
&\qquad+[(C-I_p)^{\perp}d_2]'(C-I_p)^{\perp}d_2\\
&\quad=E(\|CX+(C-I_p)d_1-\mu\|^2)+[(C-I_p)^{\perp}d_2]^{\tau}(C-I_p)^{\perp}d_2.
\end{aligned}
$$

若 $(C-I_p)^{\perp}d_2\neq0$, 显然 $CX+(C-I_p)d_1\prec CX+d$, 故 iii) 的条件是必要的.

令 $Y=Z+d_1$, 则 $Y\sim N(\mu+d_1,\ A)$, 且

$CX+(C-I_p)d_1 \overset{\text{ad. } I}{\sim} \mu \Leftrightarrow CY \overset{\text{ad. } I}{\sim} \mu+d_1 \Leftrightarrow$ i)、ii) 成立. 至此定理得证.

根据定理 4.4 可知, 定理 4.8、4.7 对于损失函数 $\|\mu-d\|^2$ 换为 $(\mu-d)'B(\mu-d)$, $B>0$. 那么结论仍然成立. 若 $X\sim N(\mu,\sigma^2A)$, $A>0$ 已知, μ、σ^2 未知. 如补充 $\frac{s^2}{\sigma^2}\sim\chi_n^2$, 那么定理 4.7、4.8 仍然成立. 充分性利用 $\sigma=1$ 时是容许的. 从而不次于它的估计必与它几乎相等. 故在任意 σ 时, 也是容许的. 必要性证明与 $\sigma=1$ 时相同. 只注意此时要用到 $\frac{s^2}{\sigma^2}\sim\chi_n^2$ 及定理 4.4.

最后, 我们还要指出一点: 在定理 4.2、定理 4.4～定理 4.6 中如果将损失函数改为 $(\mu-d)'B(\mu-d)$, $(\mu-d)'B(\mu-d)/\sigma^2$, $B>0$ 只要把估计中 $X'A^{-2}X$ 换为 $X'A^{-1}B^{-1}A^{-1}X$, A^{-1} 换为 $B^{-1}A^{-1}$. 结论仍然成立.

定理 4.6 还可推广到位置刻度参数族具有密度函数 $\sigma^{-p}f\left(\frac{x-\theta}{\sigma}\right)$, $\theta\in R^p$ 上. 证法基本相同.

§5 $\overline{X}$是位置参数容许估计的正态特性

若随机变量 X 服从非退化分布 $F(x-\theta)(-\infty<\theta<\infty)$, x_1, x_2, $\cdots$, x_n 是 X 的简单抽样，众所周知，θ 与 $E_\theta X$ 仅仅相差一个常数，$\theta=E_\theta(X)-E_0(X)$. 不失一般性，可以假设 $E_0(X)=0$. 此时用样本均值 $\overline{X}$ 估计 $\theta=E_\theta X$ 是合理的，且已证明，在平方损失下，$\overline{X}$ 就是正态均值 θ 的容许估计. 人们要问，对其它转移分布族，$\overline{X}$ 是否为容许估计呢？Kagan, Linnik 和 Rao[20] 研究了这个问题. 回答是：仅在正态分布下才是对的. 本节将介绍这方面的工作.

因为 $\overline{X}$ 是平方损失下对平移群的同变估计，如果它是容许的，必然是最优同变估计，由第七章中所述可知，最优同变估计可表为

$$\overline{X}-E_0(\bar{x}\,|\,x_2-x_1,\ \cdots,\ x_n-x_1).$$

而且它是唯一的. 这就要求 $E_0(\overline{X}\,|\,x_2-x_1,\ \cdots,\ x_n-x_1)=0$. 解决此问题的中心就是找符合此要求的分布.

引理 5.1 若 $\overline{Y}$ 是 K 维随机向量，其分布为 P, X 是随机变量，EX 存在有限. 则 $E(X\,|\,\overline{Y})=EX$ *a. s.* $P \Leftrightarrow$

$$EXe^{it'y}=EXEe^{it'y}\quad (t\in R^K). \tag{5.1}$$

证 注意对 $t\in R^k$

$$EXe^{it'y}=EE(Xe^{it'y}\,|\,Y)=E(E(X\,|\,Y)e^{it'y}).$$

必要性：由于设 $E(X\,|\,Y)=EX$ a. s $[P]$, 所以

$$EXe^{it'y}=EXEe^{it'y}\quad (t\in R^K).$$

充分性：先设 $EX\neq 0$, 故

$$Q(A)\triangleq\int_A\frac{E(X\,|\,Y)}{EX}\,dP(Y)$$

是广义测度. (5.1)式即

$$\int e^{it'y}\,dP=\int e^{it'y}\left(\frac{E(X\,|\,Y)}{EX}\right)dP=\int e^{it'y}\,dQ$$

由傅里叶变换的唯一性可知，对任何 $A\in R^k$ 上 Borel 集

$$\int_A dQ=\int_A \frac{E(X|Y)}{EX}dP=\int_A dP.$$

从而知 $E(X|Y)=EX$ a. s P.

若 $EX=0$, 则对 $b\neq 0$, 由(5.1)知

$$E(X+b)e^{it'y}=bEe^{it'y}=E(X+b)Ee^{it'y}.$$

再利用已得的结论,

$$E(X+b|Y)=E(X|Y)+b=E(X+b)$$
$$=EX+b=b \text{ a. s } P.$$

从而仍得 $E(X|Y)=0$ a. s P.

引理 5.2 若随机变量 $X\sim F(x-\theta)$, $E_\theta X^2<\infty$, $E_0X=0$, $x_1, x_2, \cdots, x_n (n\geqslant 3)$ 是 X 的简单抽样. 如果

$$E_0(\overline{X}|x_2-x_1, \cdots, x_n-x_1)=0, \tag{5.2}$$

则 $F(x)$ 为正态分布 $N(0, \sigma^2)$ ($\sigma^2>0$ 已知).

证 记 $F(x)$ 的特征函数($c. f$)为 $\varphi(t)$, 根据引理 5.1,

$$0=E_0\left\{\overline{X}\exp\left[i\left(\sum_{j=2}^n t_jX_j-\sum_{j=2}^n t_jX_1\right)\right]\right\}$$
$$=\sum_{r=1}^n E\left\{X_r\exp\left[i\sum_{j=2}^n t_j(X_j-X_1)\right]\right\}\Big/n$$
$$=-\frac{i}{n}\left(\sum_{r=2}^n\left\{\prod_{j=2}^n\varphi(t_j)\right\}\varphi'(t_r)\varphi^{-1}(t_r)\varphi\left(-\sum_{j=2}^n t_j\right)\right.$$
$$\left.+\prod_{j=2}^n\varphi(t_j)\varphi'\left(-\sum_{j=2}^n t_j\right)\right).$$

因为 $\varphi(t)$ 是 c. f, 故存在 δ, 使当 $|t_j|\leqslant\delta (j=2, \cdots, n)$ 时,

$$|\varphi(t_j)|\geqslant\frac{1}{2}, \quad \left|\varphi\left(-\sum_{j=2}^n t_j\right)\right|\geqslant\frac{1}{2}.$$

从而将上式化为

$$\sum_{j=2}^n\varphi'(t_j)\varphi^{-1}(t_j)+\varphi'\left(-\sum_{j=2}^n t_j\right)\left(\varphi\left(-\sum_{j=2}^n t_j\right)\right)^{-1}=0.$$

令 $\psi(-t)=\varphi'(t)(\varphi(t))^{-1}$, $t_3=t_4=\cdots=t_n$, 则上式化作

$$\psi(-t_2)+(n-2)\psi(-t_3)+\psi(t_2+(n-2)t_3)=0,$$
$$|t_2|<\delta, \quad |t_3|<\delta. \tag{5.3}$$

将上式左边乘以 $x-t_2$, 再将它从 0 到 x 积分, 当 $|x|<\delta$, $|t_3|<\delta$

时，则有

$$0=\int_0^x (x-t_2)\left[\psi(t_2+(n-2)t_3)+\psi(-t_2)+(n-2)\psi(-t_3)\right]dt_2$$

$$=\int_{(n-2)t_3}^{x+(n-2)t_3}(x+(n-2)t_3-u)\psi(u)du$$

$$+\psi_1(x)+(n-2)\psi(-t_3)x^2/2.$$

由题设 $E_0X^2<\infty$，知 $\varphi''(t)$ 存在，故 $\psi(t)$ 可微，将上式先对 t_3 微分，再对 x 微分，则得

$$-(n-2)\psi((n-2)t_3)+(n-2)\psi(x+(n-2)t_3)$$

$$-(n-2)x\psi'(-t_3)=0.$$

令 $t_3=0$，则上式成为 $\psi(x)=\psi(0)+\psi'(0)x$，而

$$\psi(-t)=\varphi'(t)(\varphi(t))^{-1}.$$

从而解得 $\log\varphi(t)$ 为二次多项式．又因 $\varphi'(0)=E_0X=0$，$\varphi(0)=1$，故

$$\log\varphi(t)=at^2,\quad \varphi(t)=e^{at^2}.$$

又因为 $\overline{\varphi}(t)=\varphi(-t)$ 及 $|\varphi(t)|\leqslant 1$，所以 a 仅取负实数，记 $a=-\frac{\sigma^2}{2}$，则 $\varphi(t)=e^{-\frac{\sigma^2+t^2}{2}}$，$|t|<\delta$．由此 $\varphi(t)$ 可在 $|t|<\delta$ 展成无穷级数．因此，特征函数 $\varphi(t)$ 由它在 $|t|<\delta$ 内唯一确定，即 $\varphi(t)=\exp(-\sigma^2t^2/2)$，$|t|<\infty$．这就说明 $X\sim N(0,\ \sigma^2)$．从而证得了引理．

定理 5.1 若 $X\sim F(x-\theta)$，F 为非退化，$E_0X=0$，$E_0X^2<\infty$，$x_1,\ \cdots,\ x_n$ $(n\geqslant 3)$ 是 X 的简单抽样．在平方损失下，样本均值 $\overline{X}$ 为 θ 的容许估计的充要条件是 $F(x)$ 为正态分布 $N(0,\ \sigma^2)$．

证 充分性已在本章§2中证明了．现证必要性：由 $\overline{X}$ 是 θ 的容许估计，必是 θ 的最优同变估计，从而(5.2)成立，由引理 5.2 知 $X\sim N(0,\ \sigma^2)$．

注 1 $n\geqslant 3$ 是必要的，在 $n=2$ 时，$\frac{x_1+x_2}{2}$ 也是 $\left[\theta-\frac{1}{2},\ \theta+\frac{1}{2}\right]$ 上均匀分布 θ 的容许估计．

注 2 当 X 为二维向量，$X\sim F(x-\theta)$，$E_0X=0$，$E_0\|X\|^2<\infty$．定理结论仍成立．但对于维数 $k\geqslant3$ 时，则结论成为 $\bar{X}$ 是 θ 的最优同变估计 $\Leftrightarrow F$ 为 k 维正态 $N(0, A)$．

注 3 当 $X\sim F(\sigma^{-1}(x-\theta))$，$\theta\in R^1$，$\sigma>0$，皆未知．在抽样大小 $n\geqslant6$ 时，结论仍成立(参见 A. M. Kagan and Zinger, A. A., Sankhya Ser. A, 35 (1973). 447～454)．

§6 单个位置参数最优同变估计容许性的一般理论

在这一节，我们把第三节讨论的问题推广到一般损失函数的情况．讨论转移分布族 $P_\theta=P(x-\theta)$ 位置参数 $\theta\in R^1$ 在损失函数 $L(\theta, d)=W(\theta-d)$ 之下最优同变估计的容许性．先证明在某些条件下，最优同变估计的唯一性是保证容许性的必要条件(Farrel 定理)，然后着重介绍 Brown[7, i)]的工作．

对于转移分布族，如在 § 3 中处理的那样，若 $x=(x_1, x_2, \cdots, x_k)'$，则 $Y=(x_2-x_1, \cdots, x_k-x_1)$ 是转移群下的最大不变量．Y 的分布记为 $\nu(y)$，它与 θ 无关．令 $X_1|Y\sim P(x-\theta|y)$，$\theta\in R^1$．本节皆假设此分布对 Lebesgue 测度具有密度函数 $p(x-\theta|y)$．这样可用 (X_1, Y) 替换 X 来描述这个分布族．又如在 § 3 中指出的那样，在损失函数采用 $w(\theta-d)$ 的情况下，转移群使问题不变，它的同变估计皆可表为 $x+\varphi(y)$ 的形式．同变估计的风险函数(若存在有限)一定是常数．用 $\mathscr{D}_I$ 记所有有限风险的非随机化同变估计类．

定理 6.1(Farrel[12]) 设 $w(t)$ 为碗形函数(即当 $t_1>t_2\geqslant0$ 或 $t_1<t_2\leqslant0$ 时，$w(t_1)\geqslant w(t_2)$)．且对任意的 $a>0$，有

$$P_0(X_1\in\{t: w(t)-w(t+a)>0\}|y)>0.\ \text{a. s. } \nu. \qquad (6.1)$$

如果存在两个不同的最优同变估计 $x+\varphi_i(y)$ $(i=1, 2)$ $(\nu\{\varphi_1(Y)\neq\varphi_2(y)\}>0)$，则所有的同变估计是非容许的．

证 记 $\qquad m(y)=\min\{\varphi_1(y), \varphi_2(y)\}$，

$$M(y)=\max\{\varphi_1(y),\ \varphi_2(y)\},$$

$$\delta_m(x,\ y)=x+m(y),\quad \delta_M(x,\ y)=x+M(y),$$

$$A_+=\{y:\ \varphi_1(y)>\varphi_2(y)\},\quad A_-=\{y:\ \varphi_1(y)\leqslant\varphi_2(y)\}.$$

则 δ_m 的风险函数为(注意 $x+\varphi_i(y)$ 的最优同变性)

$$\begin{aligned}R(\theta,\ \delta_m)&=\left(\int_{A+}+\int_{A-}\right)d\nu\int_{-\infty}^{\infty}w(x+m(y))p(x|y)dx\\&=\int_{A+}d\nu\int_{-\infty}^{\infty}w(x+\varphi_2(y))p(x|y)dx\\&\quad+\int_{A-}d\nu\int_{-\infty}^{\infty}w(x+\varphi_1(y))p(x|y)dx\\&=\int d\nu\inf_{\varphi}\int_{-\infty}^{\infty}w(x+\varphi(y))p(x|y)dx.\end{aligned}\tag{6.2}$$

说明 δ_m 也是 θ 的最优同变估计，同理，δ_M 也是 θ 的最优同变估计．故不妨假定 $\varphi_1(y)\leqslant\varphi_2(y)$(当 $y\in R^{k-1}$)．

现作同变估计 δ_η: 当 $x>\eta$ 时，$\delta_\eta=\delta_m$; 当 $x\leqslant\eta$ 时，$\delta_\eta=\delta_M$. δ_η 的条件风险为

$$\begin{aligned}R(\theta,\ \delta_\eta|y)&=\int_{-\infty}^{\infty}w(\delta_\eta-\theta)p(x-\theta|y)dx\\&=\int_{-\infty}^{\eta}w(x+\varphi_2(y)-\theta)p(x-\theta|y)dx\\&\quad+\int_{\eta}^{\infty}w(x+\varphi_1(y)-\theta)p(x-\theta|y)dx\\&=\int_{-\infty}^{\eta-\theta}w(x+\varphi_2(y))p(x|y)dx\\&\quad+\int_{\eta-\theta}^{\infty}w(x+\varphi_1(y))p(x|y)dx\\&=\inf_{\varphi}\int_{-\infty}^{\infty}w(x+\varphi(y))p(x|y)dx\\&\quad+\int_{-\infty}^{\eta-\theta}\varDelta(x,\ y)p(x|y)dx,\end{aligned}\tag{6.3}$$

此处 $$\varDelta(x,\ y)=w(x+\varphi_2(y))-w(x+\varphi_1(y)).$$

由于 $\varphi_1(y)\leqslant\varphi_2(y)$(当 $y\in R^{k-1}$)及 $w(t)$ 为碗形函数知 $\varDelta(x,\ y)$ 在固定 y 时至多只有一次由负到正的变号，若存在，记此变号点为

$\xi(y)$（即对任意小的 $\varepsilon>0$，在 $(-\infty,\ \xi(y)+\varepsilon)$ 上存在两点 $x_1\leqslant\xi(y)<x_2$ 使 $\Delta(x_1,\ y)\Delta(x_2,\ y)<0$；若不存在，则令 $\xi(y)=+\infty$。由于

$$\int_{-\infty}^{\infty}\Delta(x,\ y)p(x|y)dx=0,\quad \text{a. s. } \nu. \tag{6.4}$$

故不难看出，对任意实数 b，有

$$\int_{-\infty}^{b}\Delta(x,\ y)p(x|y)dx\leqslant 0,\quad \text{a. s. } \nu. \tag{6.5}$$

根据假设(6.1)式，显然有

$$\int_{-\infty}^{\xi(y)}\Delta(x,\ y)p(x|y)dx<0,\quad \text{当 } y\in\{y: \varphi_1(y)<\varphi_2(y)\},\ \text{a. s. } \nu. \tag{6.6}$$

根据(6.4)，$|\xi(y)|<\infty$，取 $\eta=\xi(y)$，由(6.3)及(6.5)式知

$$R(\theta,\ \delta_\eta)=\int d\nu \inf_{\varphi}\int_{-\infty}^{\infty}w(x+\varphi)p(x|y)dx + \int d\nu\int_{-\infty}^{\xi(y)-\theta}\Delta(x,\ y)p(x,\ y)dx. \tag{6.7}$$

而且由(6.6)式知(6.7)式的右端第二项在 $\theta=0$ 时为负，说明 δ_η 优于一切同变估计．定理得证．

下面就固定样本的情况下证明 Brown 定理(原文包括序贯抽样情况，其思路及证明方法与固定样本情况的思路、方法相似)．

定理 6.2(Brown[7,4]) 设 δ_0 为最优同变估计，

$$R_0\triangleq R(\theta,\ \delta_0)<\infty.$$

而且 i)若 $\delta_i\in\mathscr{D}_I(i=1,\ 2,\ \cdots)$ 使得 $R(\theta,\ \delta_i)\to R_0$(当 $i\to\infty$)，则

$$\delta_i(x,\ y)\to\delta_0(x,\ y)=x+\varphi_0(y)\quad (\text{当 } i\to\infty),\quad \text{a. s. } \nu;$$

ii)

$$f(\lambda)\triangleq\sup_{\varphi}\int d\nu\int_{-\lambda}^{\lambda}[w(x+\varphi_0(y)) - w(x+\varphi(y))]p(x|y)dx=O(\lambda^{-1}), \tag{6.8}$$

$$\int_0^\infty f(\lambda)d\lambda<\infty.$$

iii) $\displaystyle\int d\nu\int_{-\infty}^{\infty}|x|w(x+\varphi_0(y))p(x|y)<\infty.$

则同变估计 $\delta_0=x+\varphi_0(y)$ 是容许的．

证 如果存在另外估计 δ 使 $R(\theta, \delta)\leqslant R_0$, 记 $\psi(x, y)=\delta(x, y)-x$, 则对任意的 $l>0$, 有

$$
\begin{aligned}
0&\leqslant\int_{-l}^{l}[R_0-R(\theta, \delta)]d\theta\\
&=\int_{-l}^{l}d\theta\int d\nu\int_{-\infty}^{\infty}\{w(x+\varphi_0(y)-\theta)\\
&\quad -w(x+\psi(x, y)-\theta)\}p(x-\theta|y)dx\\
&=\int d\nu\int_{-\infty}^{\infty}dx\int_{x-l}^{x+l}\{w(z+\varphi_0(z))-w(z+\psi(x, y))\}p(z|y)dz\\
&=\int d\nu\Big[\Big\{\int_{-l/2}^{l/2}dx\int_{-l/2}^{l/2}dz+\int_{-3l/2}^{-\frac{l}{2}}dx\int_{-l/2}^{l+x}dz+\int_{l/2}^{3l/2}dx\int_{x-l}^{l/2}dz\\
&\quad+\int_{l/2}^{\infty}dz\int_{z-l}^{z+l}dx+\int_{-\infty}^{-l/2}dz\int_{z-l}^{z+l}\Big\}\cdot\bar{w}(z, x, y)p(z|y)\Big]\\
&\triangleq R_{\text{I}}+R_{\text{II}}+R_{\text{III}}+R_{\text{IV}}+R_{\text{V}}.
\end{aligned}\tag{6.9}
$$

此处 $\quad\bar{w}(z, x, y)=w(z+\varphi_0(y))-w(z+\psi(x, y))$.

由于 $z+\varphi_0(y)$ 的最优同变性质及 $z+\psi(x, y)$ 在固定 x 时也是同变的. 因此

$$R=\int_{-\infty}^{\infty}dx\Big\{\int d\nu\int_{-\infty}^{\infty}\bar{w}(z, x, y)p(z|y)dz\Big\}\leqslant 0.\tag{6.10}$$

若我们能证明

$$\limsup_{l\to\infty}(R_{\text{II}}+R_{\text{III}}+R_{\text{IV}}+R_{\text{V}})\leqslant 0,\tag{6.11}$$

$$-\infty<\liminf_{l\to\infty}R_{\text{I}}\leqslant R.\tag{6.12}$$

则由(6.9)、(6.11)式可导出 $\liminf\limits_{l\to\infty}R_{\text{I}}\geqslant 0$, 再由(6.10)及(6.12)可导出 $R=0$. 从而得出(记 $\delta_x(z, y)\triangleq z+\psi(x, y)$)

$$R(\theta, \delta_0)=R(\theta, \delta_x(z, y)),\ \text{a. s. } L.$$

再根据假设 i)就得到对 a. s. L 的 x.

$\varphi_0(y)=\psi(x, y)$. a. s. ν 成立. 再用 Fubini 定理, 得

$$
\begin{aligned}
&\int d\nu\int|x+\varphi_0(y)-\delta(x, y)|p(x|y)dx\\
&=\int d\nu\int|\varphi_0(y)-\psi(x, y)|p(x|y)dx=0.
\end{aligned}
$$

从而推出 $\delta=\delta_0$ a. s. $P_0(x)$. 这就证得了本定理.

下面分几步来证明(6.11)与(6.12)式:

(1) $\limsup_{l\to\infty}(R_{\mathrm{V}}+R_{\mathrm{IV}})\leqslant 0$.

$$R_{\mathrm{V}}+R_{\mathrm{IV}}\leqslant\int d\nu\left\{\int_{l/2}^{\infty}dz\int_{z-l}^{z+l}dx+\int_{-\infty}^{-l/2}dz\int_{z-l}^{z+l}dx\right\}w(z+\varphi_0(y))p(z|y)$$
$$=2l\int d\nu\int_{|z|>l/2}w(z+\varphi_0(y))p(z|y)dz$$
$$\leqslant 4\int d\nu\int_{|z|>l/2}|z|w(z+\varphi_0(y))p(z|y)dz\to 0\ (当\ l\to\infty).$$

最后一步是根据假设 iii).

(2) 证明(6.12)式成立: 利用 $z+\psi(x, y)$ 在固定 x 时是同变的事实以及不等式

$$\int d\nu\int_{-\lambda}^{\lambda}\overline{w}(z, x, y)p(z|y)dz$$
$$\leqslant\sup_{\varphi}\int d\nu\int_{-\lambda}^{\lambda}[w(z+\varphi_0(y))-w(z+\varphi(y))]p(z|y)dz$$
$$=f(\lambda). \tag{6.13}$$

加上假设 ii) 与 iii) 就有常数 $C_1>0$, 使得

$$R_{\mathrm{II}}=\int d\nu\int_{-3l/2}^{-l/2}dx\int_{-l/2}^{l+x}\overline{w}(z, x, y)p(z|y)dz$$
$$=\int d\nu\int_{-l/2}^{l/2}dx\int_{-l/2}^{x}\overline{w}(z, x-l, y)p(z|y)dz$$
$$=\int d\nu\left\{\int_{-l/2}^{0}dx\int_{-l/2}^{x}dz+\int_{0}^{l/2}dx\int_{-l/2}^{-x}dz+\int_{0}^{l/2}dx\int_{-x}^{x}dz\right\}\overline{w}(z, x-l, y)p(z|y)$$
$$=\int d\nu\left\{\int_{-l/2}^{0}dz\int_{z}^{0}dx+\int_{-l/2}^{0}dz\int_{0}^{-z}dx+\int_{0}^{l/2}dx\int_{-x}^{x}dz\right\}\overline{w}(z, x-l, y)p(z|y)$$
$$\leqslant 2\int d\nu\int_{-l/2}^{0}|z|w(z+\varphi_0(y))p(z|y)dz+\int_0^{\infty}f(x)dx\leqslant C_1\quad(与\ l\ 无关的常数).$$

同理可证 $R_{\text{III}}\leqslant C_1$，再结合(1)的结论及(6.9)式可知：存在 $c_2>0$ 使 $R_{\text{I}}\geqslant -c_2$，对任意 $l>0$ 成立. 从而可以使用 Fatou 引理，对确定的 $k>0$ 及一串 $\lambda_i\to\infty$（当 $i\to\infty$）有下式成立：

$$\begin{aligned}
&\liminf_{i\to\infty}\int_{-\lambda_i}^{\lambda_i}dx\int d\nu\int_{-\lambda_i}^{\lambda_i}\bar{w}(z,\ x,\ y)p(z|y)dz\\
&\leqslant\liminf_{i\to\infty}\Big[\int_{-\lambda_k}^{\lambda_k}dx\int d\nu\int_{-\lambda_i}^{\lambda_i}\bar{w}(z,\ x,\ y)p(z|y)dz\\
&\qquad+\Big(\int_{-\lambda_i}^{-\lambda_k}+\int_{\lambda_k}^{\lambda_i}\Big)dx\sup_x\int d\nu\int_{-\lambda_i}^{\lambda_i}\bar{w}(z,\ x,\ y)p(z|y)dz\Big]\\
&\leqslant\liminf_{i\to\infty}\Big[\int_{-\lambda_k}^{\lambda_k}dx\int d\nu\int_{-\lambda_i}^{\lambda_i}\bar{w}(z,\ x,\ y)p(z|y)dz+2f(\lambda_i)\lambda_i\Big]\\
&=\int_{-\lambda_k}^{\lambda_k}dx\int d\nu\int_{-\infty}^{\infty}\bar{w}(z,\ x,\ y)p(z|y)dz\to R(\text{当 } k\to\infty).
\end{aligned}$$

这就证明了(6.12).

(3) 对任意给定的 $a>0$，正可测函数 $\alpha(y)$，定义

$$\begin{aligned}S(l)=\{(x,\ y):-l-a<x<-l+a,\\ |\psi(x,\ y)-\varphi_0(y)|>\alpha(y)\},\end{aligned}$$

则

$$\int_{S(l)}dx\,d\nu\to 0\ \ (\text{当 } l\to\infty). \tag{6.14}$$

证 对任意给定的 $\beta_1>0$，记

$$T(l)=\{x:\nu(y:(x,\ y)\in S(l))\}>\beta_1/4a\},\ \ T=\bigcup_{l>0}T(l).$$

从 $T(l)$、T 的定义可看到不存在 $\{x_i\}$ 使 $x_i\in T$，且 $\psi(x_i,\ y)\to\varphi_0(y)\ (i\to\infty)$. 于是由假设 i)知：存在 $\beta_2>0$，使得当 $x\in T$ 时，有

$$\begin{aligned}&R(\theta,\ z+\varphi_0(y))-R(\theta,\ z+\psi(x,\ y))\\ &=\int d\nu\int\bar{w}(z,\ x,\ y)p(z|y)dz\leqslant-\beta_2.\end{aligned}$$

由上式和(6.12)式即知

$$\begin{aligned}&-\infty<-c_2\leqslant R\\ &=\int_{-\infty}^{\infty}dx\int d\nu\int\bar{w}(z,\ x,\ y)p(z|y)dz\leqslant-\beta_2\int_T dx.\end{aligned}$$

从而 $\int_T dx\leqslant c_2/\beta_2<\infty$，故存在 l_0，当 $l>l_0$ 时，$\int_{T(l)}dx\leqslant\beta_1/2$，

从而

$$\int_{S(l)} d\nu\, dx \leqslant \int_{T(l)} dx + \int_{-l-a}^{-l+a} \beta_1 (4a)^{-1} dx \leqslant \beta_1.$$

而 β_1 是任意选取的，故(6.14)成立.

(4) $\limsup_{l\to\infty} R_{\mathrm{II}} \leqslant 0$, $\limsup_{l\to\infty} R_{\mathrm{III}} \leqslant 0$,

$$R_{\mathrm{II}} = \int d\nu \Big[\int_{-l-a}^{-l+a} dx \int_{l/2}^{l+x} dz + \int_{-3l/2}^{-l-a} dx \int_{-l/2}^{l+x} dz + \int_{-l+a}^{-l/2} dx \int_{-l-x}^{l+x} dz$$

$$+ \int_{-l+a}^{-l/2} dx \int_{-l/2}^{-l-x} dz\Big] \overline{w}(z;\ x,\ y) p(z|y), \qquad (6.15)$$

$$\{(x,\ y): -l-a < x < -l+a\}$$

$$= \{(x,\ y): -l-a < x < -l+a,\ (x,\ y) \bar{\in} S(l)\} \cup S(l).$$

由于(6.12)式中的 R 存在有限及 $\lim_{l\to\infty}\int_{S(l)} dx\, d\nu = 0$, 故

$$\lim_{l\to\infty} \int_{S(l)} d\nu\, dx \int_{-l/2}^{l+x} \overline{w}(z;\ x,\ y) p(z|y) dz = 0. \qquad (6.16)$$

因为 $0 \leqslant R_0 < \infty$ 可知对于任意的 $\varepsilon > 0$, 存在 $b > 0$ 及 $l_0 > 0$, 使得

$$\int_{|\varphi_0| > b} d\nu \int w(z + \varphi_0(y)) p(z|y) dz < \varepsilon/4a. \qquad (6.17)$$

$$\int d\nu \int_{z < -l_0/2} w(z + \varphi_0(y)) p(z|y) dz$$

$$= \int_{z < -l_0/2} dz \int w(z + \varphi_0(y)) p(z|y) d\nu < \frac{\varepsilon}{4a}.$$

由此及 Luzin 定理，对任意的 $\varepsilon > 0$, 存在任意的有界闭子集 $K(y)$ 及适当小的 $\alpha > 0$, 使得

$$\int_{z+c \bar{\in} K(y)} w(z + \varphi_0(y)) p(z|y) dz < \frac{\varepsilon}{16a} \quad (当\ |c| < \alpha),$$

$$|w(z + \varphi_0(y)) - w(z + \varphi_0(y) + \eta)| < \frac{\varepsilon}{8a}$$

$$\left(当\ z,\ z+\eta \in K(y) \cap \left[\frac{l_0}{2},\ a\right],\ |\eta| \leqslant \alpha\right)$$

由此，在 $|\psi(x,\ y) - \varphi_0(y)| \leqslant \alpha$ 及 $l > l_0$ 下有

$$\int_{|\varphi_0|<b,-l+a<x<-l-a,|\varphi_0-\psi|<\alpha} d\nu\, dx \int_{-l/2}^{l+x} \bar{w}(z,\ x,\ y) p(z|y) dz$$

$$\leqslant 2a \int_{z<-\frac{l_0}{2}} w(z+\varphi_0(y)) p(z|y) dz\, d\nu$$

$$+\int_{|\varphi_0|<b,-l+a<x<-l-a,|\varphi_0-\psi|<\alpha} d\nu\, dx \int_{-l_0/2}^{l+x} \bar{w}(z,\ x,\ y) p(z|y) dz$$

$$\leqslant \frac{\varepsilon}{2}+\frac{2a\varepsilon}{8a}\int_{|\varphi_0|<b} d\nu+\frac{2\varepsilon}{16a}\cdot 2a\int_{|\varphi_0|<b} d\nu \leqslant \varepsilon. \tag{6.18}$$

综合 (6.16)～(6.18) 及 $S(L)=\{(x,\ y),\ -l-a<x<-l+a,\ |\psi(x,\ y)-\varphi_0(y)|>\alpha\}$

$$\int d\nu \int_{-l-a}^{-l+a} dx \int_{-l/2}^{l+x} \bar{w}(z;\ x,\ y) p(z|y) dz$$

$$\leqslant \int_{S(l)} dx\, d\nu \int_{-l/2}^{l+x} w(z+\varphi_0(y)) p(z|y) dz$$

$$+\int_{|\varphi_0|>b} d\nu \int_{-l-a}^{-l+a} dx \int w(z+\varphi_0(y)) \cdot p(z|y) dz$$

$$+\int_{|\varphi_0|<b,-l-a<x<-l+a,(x,y)\bar{\in} S(l)} dx\, d\nu$$

$$\times \int_{-l/2}^{l+x} \bar{w}(z;\ x,\ y) p(z|y) dz. \tag{6.19}$$

上式当 $l\to\infty$ 时，它的上极限小于或等于 $\frac{3\varepsilon}{2}$. 由于 ε 的任意性，即知

$$\limsup_{l\to\infty} \int_{-l-a}^{-l+a} dx \int d\nu \int_{-l/2}^{l+x} \bar{w}(z;\ x,\ y) p(z|y) dz \leqslant 0. \tag{6.20}$$

然后我们再导出下述不等式：

$$\int d\nu \int_{-3l/2}^{-l-a} dx \int_{-l/2}^{l+x} \bar{w}(z;\ x,\ y) p(z|y) dz$$

$$=\int d\nu \int_{-l/2}^{-a} p(z|y) dz \int_{z-l}^{-a-l} \bar{w}(z;\ x,\ y) dx$$

$$\leqslant \int d\nu \int_{-l/2}^{-a} p(z|y) dz \int_{z-l}^{-a-l} w(z+\varphi_0(y)) dx$$

$$\leqslant \int d\nu \int_{-l/2}^{-a} (a+|z|) w(z+\varphi_0(y)) p(z|y) dz$$

$$\leqslant 2\int d\nu \int_{-\infty}^{-a} |z| w(z+\varphi_0(y)) p(z|y) dz \to 0\ (\text{当 } a\to\infty). \tag{6.21}$$

$$\int d\nu \int_a^{l/2} dx \int_{-x}^{x} \bar{w}(z,\ x-l,\ y) p(z|y) dz$$

$$\leqslant \int_a^{l/2} dx \sup_\varphi \int d\nu \int_{-x}^{x} [w(z+\varphi_0(y)) - w(z+\psi)] p(z|y) dz$$

$$\leqslant \int_a^{\infty} dx \sup_\psi \int d\nu \int_{-x}^{x} [w(z+\varphi_0(y)) - w(z+\psi)] p(z|y) dz$$

$$\to 0 \quad (当\ a\to\infty).\ (由假设\ \mathrm{ii}) \tag{6.22}$$

$$\int d\nu \int_{l+a}^{-l/2} dx \int_{-l/2}^{-x-l} \bar{w}(z;\ x,\ y) p(z|y) dz$$

$$= \int d\nu \int_{-l/2}^{-a} dz \int_{a-l}^{-z-l} \bar{w}(z;\ x,\ y) p(z|y) dx$$

$$\leqslant \int d\nu \int_{-\infty}^{-a} (a+|z|) \bar{w}(z+\varphi_0(y)) p(z|y) dz$$

$$\leqslant 2\int_{-\infty}^{-a} dz \int |z| w(z+\varphi_0(y)) d\nu \to 0 \quad (当\ a\to\infty). \tag{6.23}$$

将(6.20)至(6.23)代入(6.15)式，即得

$$\limsup_{l\to\infty} R_{\mathrm{II}} \leqslant 0.$$

对于R_{III}，证明是相同的，综合(1)、(4)的结论即得(6.11)式。由此全部证明了定理。

为了使得定理便于使用，有必要对定理的条件进行一番分析，以找到方便使用的充分条件。下面作一简单扼要的讨论：

推论 1 若$w(t)$是碗形函数，则假设 i)成立的充要条件是最优同变估计唯一存在风险有限函数(指 a. s. P)。

证 必要性是显然的，现证充分性：设 δ_0 是最优同变估计，且有同变估计序列 $\{\delta_k = x+\varphi_k(y),\ k=1,\ 2,\ \cdots\}$ 使 $R(\theta,\ \delta_k)\to R(\theta,\ \delta_0)<\infty$ （当 $k\to\infty$）。记 $R(\delta|y)=\int w(z+\varphi(y))p(z|y)dz$。则上式化为

$$\int [R(\delta_k|y) - R(\delta_0|y)] d\nu(y) \to 0 \quad (k\to\infty).$$

由于 $R(\delta_k|y)\geqslant R(\delta_0|y)$ a. s. ν，故 $R(\delta_k|y)\to R(\delta_0|y)$ 依 ν 测度。如果 $\varphi_k\to\varphi_0$ a. s ν不成立，则必有

$$0<\nu\{y:\ \limsup_{k\to\infty}\varphi_k(y)\neq\varphi_0(y)\}$$

或

$$0<\nu\{y:\ \liminf_{k\to\infty}\varphi_k(y)\neq\varphi_0(y)\}. \qquad (6.24)$$

根据 Luzin 定理对任意 b, 可证 $\int_{|z|<b} w(z+a)p(z|y)dz$ 是 a 的连续函数, 利用 $w(t)$ 为碗形函数, 当 $a\to\pm\infty$ 时, 可在积分号下取极限. 根据这两点可证明

$$\limsup_{k\to\infty} R(\delta_k|y)\geqslant\int w(z+\limsup_{k\to\infty}\varphi_k(y))p(z|y)\geqslant R(\delta_0|y);$$

$$\liminf_{k\to\infty} R(\delta_k|y)\geqslant\int w(z+\liminf_{k\to\infty}\varphi_k(y))p(z|y)\geqslant R(\delta_0|y).$$

如果(6.24)成立, 则与 $\lim_{k\to\infty}R(\delta_k|y)=R(\delta_0|y)$ (依测度 ν) 及假设最优同变估计的唯一性相矛盾. 故推论 1 成立.

推论 2 若推论 1 关于 $w(t)$ 的假设成立, 且最优同变估计唯一存在. 且

$$\int d\nu\int|x-\varphi_0|\sup\{w(2x),\ w(-2x)\}P(x-\varphi_0(y)|y)dx<\infty.$$

则假设 ii) 与 iii) 成立.

证 iii) 成立是显然的. 在此只要证明 ii) 成立. 因为 $x+\varphi_0(y)$ 是最优同变估计且唯一, 所以对于任意的 $\varphi(y)$, 有

$$\int d\nu\int_{-\infty}^{\infty}[w(x+\varphi_0(y))-w(x+\varphi(y))]p(x|y)dx\leqslant 0$$

或

$$\int d\nu\int_{-\lambda}^{\lambda}[w(x+\varphi_0)-w(x+\varphi)]p(x|y)dx$$

$$\leqslant-\int d\nu\int_{|x|>\lambda}[w(x+\varphi_0)-w(x+\varphi)]p(x|y)dx. \qquad (6.25)$$

为了讨论方便, 不妨设 $\varphi_0(y)=0$. 我们作

$$\tilde{\varphi}(y)=\begin{cases}-\lambda, & 当\ \ \varphi<-\lambda;\\ \varphi(y), & 当\ \ |\varphi|\leqslant\lambda;\\ \lambda, & 当\ \ \varphi>\lambda.\end{cases}$$

从碗形函数易知: 在 $|z|\leqslant\lambda$ 时, $w(z+\varphi)\geqslant w(z+\tilde{\varphi})$. 从而

$$\sup_{\varphi}\int d\nu\int_{-\lambda}^{\lambda}[w(z)-w(z+\varphi)]p(z|y)dz$$
$$=\sup_{|\varphi|<\lambda}\int d\nu\int_{-\lambda}^{\lambda}[w(z)-w(z+\varphi)]p(z|y)dz$$
$$\triangleq\sup_{|\varphi|<\lambda}I(\lambda,\ \varphi).$$

根据(6.25)式及积分次序变换，有

$$\int_0^{\infty}\sup_{|\varphi|<\lambda}I(\lambda,\ \varphi)d\lambda$$
$$\leqslant\int_0^{\infty}d\lambda\left\{\sup_{|\varphi|<\lambda}\int d\nu\int_{|z|>\lambda}[w(z+\varphi)-w(z)]p(z|y)dz\right\}$$
$$\leqslant\int_0^{\infty}d\lambda\int d\nu\int_{\lambda}^{\infty}\sup[w(2z),\ w(-2z)][p(z|y)+p(-z|y)]dz$$
$$\leqslant\int d\nu\int|z|\sup[w(2z),\ w(-2z)]p(z|y)dz<\infty.$$

由本假设即知假设 ii) 成立.

由这两个推论立即可得到如下两个结论:

推论 3 若 $w(t)=|t|^k$, 在 $k>1$ 时，存在同变估计 $\delta(x)$，使得 $E_0|\delta(x)|^{k+1}<\infty$; 当 $k=1$ 时，假设 $P(\cdot|y)$ 的中位数 $\varphi(y)$ 唯一(a. s. ν), 则最优同变估计是容许的 minimax 估计.

证 因为 $|t|^k$ 在 $k>1$ 是严凸的，而

$$\delta(x)=x+p\varphi_1(y)+q\varphi_2(y)$$

也是同变的. 这说明最优同变估计是唯一的.

推论 4 若 p 维随机向量 X 具有密度函数 $\sigma^{-p}p\left(\frac{x}{\sigma}\right)$(当 $x_i>0,\ i=1,\ 2,\ \cdots,\ p$); 其它为零. 损失函数采用 $\left|\frac{d}{\sigma}-1\right|^k(k\geqslant1)$, 并假设存在刻度变换群的同变估计 $\delta(x)$，使得 $E_1\delta^k(x)|\log\delta(x)|<\infty$, 在 $k=1$ 时，最优同变估计唯一. 则最优同变估计是容许的 minimax 估计.

证 令 $\ln x_i=y_i,\ \theta=\log\sigma,\ \tilde{d}=\log d$, 则

$$y=(y_1,\ \cdots,\ y_p)^{\tau}\sim\prod_1^p e^{y_i-\theta}p(e^{y_1-\theta},\ \cdots,\ e^{y_p-\theta}).$$

损失函数化为$|\exp(\bar{d}-\theta)-1|^k$下的估计$\theta$, 变成为位置参数估计问题. 利用推论 1、2 及定理 6.2 即得证.

§7 关于二维位置参数最优同变估计的容许性

对于多个参数的转移分布族,参数的最优同变估计的容许性,并非想像那样,是一维的自然推广. 在参数多于 2 时,象正态分布均值的最优同变估计都不是容许的,其它分布族就更难设想了. 但对于位置参数只有两个时, 在一定条件下, 它的最优同变估计是容许的. 这一节就来讨论这个问题, 介绍 Brown 和 Fox[8] 的工作.

设 X 是二维随机向量, Y 为多维随机向量, $X|Y$ 有关于 L 的概率密度 $p(x-\theta|y)$, $\theta\in R^2$, $Y\sim\nu(y)$与θ无关. 损失函数采用 $L(\theta, d)=w(\theta-d)$. 在 R^2 空间有转移群 $\mathscr{G}=\{g: g\in R^2\}$, 当 $g\in\mathscr{G}$ 时, 有 $gx=x+g$, $gy=y$, $\mathscr{G}$ 在参数空间与行动空间(均为 R^2)引出与 $\mathscr{G}$ 同构的加法群 $\overline{\mathscr{G}}$, $\mathscr{G}^*$, 这样, 我们的估计问题是不变判决问题.

θ 的一个非随机化估计 $\delta(x, y)$ **称为同变的**, 如果对任意的 $g\in\mathscr{G}$, 皆有

$$\delta(gx, gy)=\delta(gx, y)=g^*\delta(x, y).$$

取 $g=-x$, 立即得到同变估计形式可表为

$$\delta(x, y)=\delta(0, y)+x\triangleq x+\varphi(y). \tag{7.1}$$

假设 1. 存在基本唯一的最优同变估计 $\delta_0=x+\varphi_0(y)$ 具有有限风险,即

$$R_0=R(\theta,\delta_0)=\inf_{\varphi}R(\theta, x+\varphi(y))<\infty.$$

不失一般性,假设 $\varphi_0(y)=0$(否则,取 $\tilde{x}=x+\varphi_0(y)$, 样本空间换为 $\tilde{x}$, y 变量空间, $p(x|y)$换为 $p(\tilde{x}|y)$, 此时 $\delta_0=\tilde{x}$).

假设 2.

$$\int d\nu\int\|x\|^2w(x)p(x|y)dx<\infty. \tag{7.2}$$

假设 3. 若存在估计 δ, 使得对于 $\forall\theta\in R^2$, $R(\theta,\ \delta)\leqslant R_0$, 则存在一串估计 δ_l, 具有

3.1) $\delta_l\to\delta$ (当 $l\to\infty$) a. s. L;

3.2) $R(\theta,\ \delta_l)\leqslant R_0$ (当 $\|\theta\|\leqslant l$);

3.3) 对于 $l\leqslant m$, $\|x\|<3l/2$ 时,

$$\int w(\delta_l-x+z)p(z|y)dz\leqslant\int w(\delta_m-x+z)p(z|y)dz; \quad (7.3)$$

3.4) $$\lim_{l\to\infty}\int d\nu\int_{\|x\|<l/2}dx\int_{\|\theta\|>l}w(\delta_l-\theta)p(x-\theta|y)d\theta$$

$$=\lim_{l\to\infty}\int d\nu\int_{\|x\|<l/2}dx\int_{\|x-z\|>l}w(\delta_l-x+z)p(z|y)dz=0.$$

假设 4. 存在非增函数 $K(v)$, 由 $(0,\ \infty)\to(0,\ \infty)$ 使得

$$\int_0^\infty K(v)dv=c_0<\infty, \quad (7.4)$$

且对任意同变估计 δ_I, 下面两个不等式成立:

$$\int d\nu\int_{\|x\|<v}[w(x)-w(\delta_I(x,\ y))]p(x|y)dx$$

$$\leqslant K(v)\left\{\int d\nu\int[w(\delta_I)-w(x)]p(x|y)dx\right\}^{1/2}; \quad (7.5)$$

$$\int d\nu\int_{\|x\|>v}[w(x)-w(\delta_I)]^+p(x|y)dx$$

$$\leqslant\left\{\int d\nu\int[w(\delta_I)-w(x)]p(x|y)dx\right\}^{1/2}K(v). \quad (7.6)$$

定理 7.1 若假设 1~4 成立, 则 $\delta_0(x,\ y)=x$ 是容许的.

证 若存在 δ 使得

$$R(\theta,\ \delta)\leqslant R_0=R(\theta,\ \delta_0) \quad (\text{当 } \theta\in R^2).$$

由假设 3 知, 存在 δ_l, 使得

$$0\leqslant\int_{\|\theta\|<l}[R_0-R(\theta,\ \delta_l)]d\theta$$

$$=\int d\nu\int dx\int_{\|\theta\|<l}[w(x-\theta)-w(\delta_l-\theta)]p(x-\theta|y)d\theta$$

$$=\int d\nu\int dx\int_{\|x-z\|<l}[w(z)-w(\delta_l-x+z)]p(z|y)dz$$

$$=\left\{\int d\nu\int_{\|x\|<l/2}dx\int_{\|x-z\|<l}dz+\int d\nu\int_{l/2<\|x\|<3l/2}dx\int_{\|x-z\|<l}dz\right.$$

$$\left.+\int d\nu\int_{\|x\|>3l/2}dx\int_{\|x-z\|<l}dz\right\}[w(z)-w(\delta_l-x+z)]p(z|y)$$

$$\triangleq I_1(l)+I_2(l)+I_3(l). \tag{7.7}$$

根据假设 2，易知

$$I_3(l)\leqslant\int d\nu\int_{\|x\|>3l/2}dx\int_{\|x-z\|<l}w(z)p(z|y)dz$$

$$\leqslant\int d\nu\int_{\|z\|>l/2,\|x-z\|<l}w(z)p(z|y)dx\,dz$$

$$=\pi l^2\int d\nu\int_{\|z\|>l/2}w(z)p(z|y)dz$$

$$\leqslant 4\pi\int d\nu\int_{\|z\|>l/2}\|z\|^2w(z)p(z|y)dz\to 0\quad(\text{当 } l\to\infty). \tag{7.8}$$

$$h(l)\triangleq\int d\nu\int_{\|x\|<l/2}dx\int_{\|x-z\|>l}[w(z)-w(\delta_l-x+z)]p(z|y)dz$$

$$\leqslant\int d\nu\int_{\|x\|<l/2}dx\int_{\|z\|>l/2}w(z)p(z|y)dz$$

$$=4^{-1}\pi l^2\int d\nu\int_{\|z\|>l/2}w(z)p(z|y)dz$$

$$\leqslant\pi\int d\nu\int_{\|z\|>l/2}\|z\|^2w(z)p(z|y)dz\to 0\quad(\text{当 } l\to\infty). \tag{7.9}$$

另一方面，由假设 3 的 3.4) 知：

$$h(l)\geqslant-\int d\nu\int_{\|x\|<l/2}dx\int_{\|x-z\|>l}w(\delta_l-x+z)p(z|y)dz$$

$$\to 0\quad(\text{当 } l\to\infty). \tag{7.10}$$

因此 $h(l)=o(1)$. 而

$$I_2(l)\leqslant\int d\nu\int_{l/2<\|x\|<3l/2}dx\int w(z)p(z|y)dz=2\pi R_0l^2. \tag{7.11}$$

由于在 (z, y) 变量空间中，z 是最优同变估计，而当固定 x 时，$\delta(x, y)-x+z$ 也是同变估计，故

$$\int[w(\delta_l-x+z)-w(z)]p(z|y)dz\geqslant 0 \tag{7.12}$$

对任意的 x、y 成立. 所以

$$I(l) \triangleq -\int d\nu \int_{\|x\|<l/2} dx \int [w(z)-w(\delta_l - x+z)] p(z|y) dz \geqslant 0. \tag{7.13}$$

由 $I(l)$ 的定义易知: 它是 l 的单调非降函数, 根据(7.7),

$$I(l) = -I_1(l) - h(l) \leqslant I_2(l) + I_3(l) - h(l).$$

再运用(7.8)～(7.13)式, 可知

$$I(l) = O(l^2), \quad I_2(l) \geqslant I(l) + o(1). \tag{7.14}$$

由集合等式

$$\begin{aligned}
&\left\{\frac{l}{2} \leqslant \|x\| < 3l/2, \|x-z\| < l\right\} \\
&\quad = \left\{\frac{l}{2} \leqslant x < l, \ \|z\| < l - \|x\|\right\} \\
&\qquad \cup \{l/2 \leqslant \|x\| < 3l/2, \ \|x-z\| < l, \ \|z\| > |l - \|x\||\},
\end{aligned}$$

$$\begin{aligned}
l^{-1}\int_l^{2l} I_2(t)dt &\leqslant l^{-1}\int_l^{2l} dt \int d\nu \Big[\int_{t/2<\|x\|<t} dx \\
&\quad \times \int_{\|z\|<t-\|x\|} \{w(z) - w(\delta_l - x+z)\} p(z|y) dz] \\
&\quad + \int_{t/2<\|x\|<3t/2} dx \int_{\|z\|>|t-\|x\||} \{w(z) - w(\delta_l - x+z)\}^+ p(z|y) dz \\
&\leqslant l^{-1}\int_l^{2l} dt \int_{t/2<\|x\|<3t/2} 2K(|t-\|x\||) \\
&\quad \times \left[\int d\nu \int (w(\delta_l - x+z) - w(z)) p(z|y) dz\right]^{1/2} dx \\
&\leqslant 2l^{-1}\int_{l/2<\|x\|<3l} dx \left[\int d\nu \int [w(\delta_{6l} - x+z) - w(z)] p(z|y) dz\right]^{1/2} \\
&\quad \times \int_{\max(l, 2\|x\|/3)}^{\min(2l, 2\|x\|)} K(|t-\|x\||) dt \\
&\leqslant c_1\left\{\int_{l/2<\|x\|<3l} dx \int d\nu \int [w(\delta_{6l} - x+z) - w(z)] p(z|y) dz\right\}^{1/2} \\
&\leqslant c_1\left\{\int_{\|x\|<3l} dx \int d\nu \int [w(\delta_{6l} - x+z) - w(z)] p(z|y) dz\right. \\
&\quad \left. - \int_{\|x\|<l/2} \int d\nu \int [w(\delta_l - x+z) - w(z)] p(z|y) dz\right\}^{1/2} \\
&= c_1[(I(6l) - I(l)]^{1/2}.
\end{aligned} \tag{7.15}$$

此处

$$2l^{-1}\left\{\int_{l/2\leqslant\|x\|<3l} dx\left[\int_{\max(l,2\|x\|/3)}^{\min(2l,2\|x\|)} K(|t-\|x\||)dt\right]^2\right\}^{1/2}$$

$$\leqslant 2(8\pi c_0^2)^{1/2}=c_1<\infty.$$

注意 c_1 为有限数，且与 l 无关(还可看到当维数 $\geqslant 3$ 时，c_1 就不是有限的了，说明 3 维以上的位置参数最优同变估计多数是非容许的)．以上在证明(7.15)中第二个不等式是利用假设 4；第三个不等式是采用积分换序及假设 3.3)；第四个不等式是利用 Schwarz 不等式；第五个不等式是用假设 3.3.

因此由(7.14)或(7.15)式知，对任意的 $l>0$，存在 l' 满足 $l\leqslant l'\leqslant 2l$，使

$$\begin{aligned} I(6l)-I(l) &\geqslant c_1^{-2}I_2^2(l')\geqslant c_1^{-2}(I(l')+o(1))^2 \\ &\geqslant c_1^{-2}(I(l)+o(1))^2. \end{aligned} \tag{7.16}$$

若 $I(l)$ 无界，将上式两边除以 $I(l)$，即知 $I(6l)/I(l)$ 也趋于无穷，即存在 l_0，当 $l\geqslant l_0$ 时，

$$I(6l)/I(l)\geqslant 6^3,\quad I(6^n l_0)\geqslant 6^{3n}I(l_0)\quad (n=1,\ 2,\ \cdots).$$

这与 $I(l)=O(l^2)$ 矛盾，故 $I(l)$ 有界．将(7.16)两边对 $l\to\infty$ 取极限，立即得出 $I(l)\to 0$（当 $l\to\infty$）．前已知 $I(l)\geqslant 0$ 且非降，故 $I(l)=0$．从 $I(l)$ 的定义(7.13)式看，注意(7.12)式成立，故可知 $\|x\|<l/2$ 时，有

$$\int[w(\delta_l-x+z)-w(z)]p(z|y)dz=0,\quad \text{a. s. } \nu\times L.$$

再由假设 1 最优同变估计的唯一性，就得到

$$\delta_l-x+z=z,\quad \text{a. s. } p(z|y)dz\,d\nu$$

在 $\|x\|<l/2$ 上 $\delta_l(x,\ y)=x$, a. s. ν．由假设 3 知

$$\delta_l(x,\ y)\to\delta(x,\ y)\quad (\text{当 } l\to\infty)\quad \text{a. s. } p(x|y)dx\,d\nu.$$

故 $\delta(x,\ y)=x$, a. s. $p(x|y)dx\,d\nu$．因此 $\delta_0(x,\ y)=x$ 是容许的．定理证完．

定理 7.2 若 $w(t)=\|t\|^2=t_1^2+t_2^2$，存在一个同变估计 δ_I 使

$$E_0(\|\delta_I\|^4|\log\|\delta_I\||^{2+\varepsilon})<\infty\quad (\varepsilon>0), \tag{7.17}$$

则最优同变估计是容许的．

证 $\|t\|^2$ 是严凸函数，可得存在唯一(a. s)的最优同变估计 δ_0（不失一般性，可假设 $\delta_0(x, y)=x$）。由此及(7.17)知定理 7.1 中的假设 1、2 成立.

若有 δ 不比 δ_0 差. 那么取

$\delta_l(x, y)=\delta$（当 $\|\delta\|\leqslant B_y(l)$），$\delta_l(x, y)=0$（当 $\|\delta\|>B_y(l)$），

这里 $B_y(l)=4R_0^{1/2}(y)+9l/2$, $R_0(y)=\int\|x\|^2 p(x|y)dx$.

根据这样定义的 δ_l 就有

$$\|\delta_l-\theta\|^2\leqslant\|\delta-\theta\|^2 \quad (\text{当 } \|\theta\|<l);$$

$$R(\theta, \delta_l)\leqslant R(\theta, \delta)\leqslant R_0 \quad (\text{当 } \|\theta\|<l); \tag{7.18}$$

$$\delta_l\to\delta, \quad \text{a. s. } p(x|y)dx\,d\nu. \tag{7.19}$$

当 $\|\delta\|\leqslant B_y(l)$ 或 $\|\delta\|>B_y(m)$（$l\leqslant m$ 时），有

$$\|\delta_l-x+z\|^2=\|\delta_m-x+z\|^2,$$

而 $\|\delta\|>B_y(l)$ 时，在 $\|x\|<3l/2$ 下，有

$$\begin{aligned}\int\|-x+z\|^2p(z|y)dz&\leqslant2\int(\|x\|^2+\|z\|^2)p(z|y)dz\\&\leqslant2\left(\left(\frac{3l}{2}\right)^2+R_0(y)\right)\leqslant[(B_y(l)-3l/2)/2]^2/2\\&\leqslant\int_{\|z\|<(B_y(l)-3l/2)/2}\|z+\delta-x\|^2p(z|y)dz\\&\leqslant\int\|z+\delta-x\|^2p(z|y)dz.\end{aligned} \tag{7.20}$$

(7.20)中的第三个不等号是利用 Tchebyshev 不等式

$$\int_{\|z\|<(B_y(l)-3l/2)/2}p(z|y)dz$$
$$\geqslant1-R_0(y)(B_y(l)-3l/2)_4^{-2}\geqslant\frac{1}{2}.$$

由(7.18)～(7.20)知，定理 7.1 中的假设 3.1)～3.3) 成立.

$$\|\delta_l-x+z\|^2\leqslant3(\|z\|^2+\|\delta_l\|^2+\|x\|^2)$$

$$\leqslant3\left(\|z\|^2+183\left(\frac{l}{2}\right)^2+32R_0(y)\right) \quad (\text{当 } \|x\|<l/2).$$

因此

$$\int d\nu\int_{\|x\|<l/2}dx\int_{\|x-z\|>l}\|\delta_l-x+z\|^2p(z|y)dz$$

$$\leqslant\int d\nu\int_{\|x\|<l/2}dx\int_{\|z\|>l/2}3\left(\|z\|^2+183\left(\frac{l}{2}\right)^2+32R_0(y)\right)p(z|y)dz$$

$$\leqslant3\pi\int d\nu\int_{\|z\|>l/2}\left(\frac{l}{2}\right)^2\left(\|z\|^2+183\left(\frac{l}{2}\right)^2+32R_0(y)\right)p(z|y)dz$$

$$\leqslant552\pi\int d\nu\int_{\|z\|>l/2}\|z\|^4p(z|y)dz$$

$$+96\pi\int R_0(y)d\nu\int_{\|z\|>l/2}\|z\|^2p(z|y)dz$$

$$\leqslant552\pi\int d\nu\int_{\|z\|>l/2}\|z\|^4p(z|y)dz$$

$$+96\pi\left(\int d\nu\int\|z\|^4p(z|y)dz\right)^{1/2}\left(\int d\nu\int_{\|z\|>l/2}\|z\|^4p(z|y)dz\right)^{1/2}$$

$$\to0\quad(\text{当 } l\to\infty).$$

上式最后一个不等式是两次利用 Schwarz 不等式，而 $\to0$ 是根据假设(7.17). 因此假设 3.3 成立.

选
$$K(t)=2\left(\int d\nu\int_{\|x\|>t}\|x\|^2p(x|y)dx\right)^{1/2}$$

$$K(t)\leqslant2t^{-1}\log^{-(1+\varepsilon/2)}t\left\{\int d\nu\int_{\|x\|>t}\|x\|^4\log^{2+\varepsilon}\|x\|p(x|y)dx\right\}^{1/2}$$

$$\leqslant\text{const}\,t^{-1}\log^{-(1+\varepsilon/2)}t\quad(\text{当 } t>1).$$

故
$$\int_0^\infty K(t)dt<\infty.$$

注意到 $\int xp(x|y)dx=0$, 故

$$\left(\int d\nu\int_{\|x\|<t}(\|x\|^2-\|x+r(y)\|^2p(x|y)dx\right)^+$$

$$=\left(\int d\nu\int_{\|x\|<t}(-2x'r(y))p(x|y)dx\right)^+$$

$$=\left(\int d\nu\int_{\|x\|\geqslant t}2x'r(y)p(x|y)dx\right)^+$$

$$\leqslant2\left(\int d\nu\int_{\|x\|\geqslant t}\|x\|^2p(x|y)dx\right)^{1/2}\left(\int d\nu\int\|r(y)\|^2p(x|y)dx\right)^{1/2}$$

$$=K(t)\left\{\int d\nu\int(\|x+r(y)\|^2-\|x\|^2)p(x|y)dx\right\}^{1/2}.$$

故(7.5)成立. 同理(7.6)式成立. 至此已验证定理7.1的全部假设都成立. 因此最优同变估计是容许的.

推论 若 $X \sim N(\mu, I_2)$, $\mu \in R^2$, 则 x 是在平方损失下 μ 的容许估计.

问题与习题

1. 令 $\{X_i\}$ 为一串相互独立的随机变量, $X_i \sim$ 二项分布 $B(n, p_i)$ $(0<p_i<1)$. 求证 $\frac{1}{m}(X_1, \cdots, X_m)'$ 是 $(p_1, \cdots, p_m)'$ 在损失函数 $\sum_{i=1}^{m}(p_i-d_i)^2$ 下的容许估计.

$\left[\text{提示: } \frac{X_i}{n}\right.$ 是在 $(p_i-d_i)^2/p_i(1-p_i)$ 损失下, 关于均匀分布的 Bayes minimax 估计. 利用定理1.5.]

2. 若随机向量 $(X_1, \cdots, X_k)$ 具有多项分布 $B(n, p_1, \cdots, p_k)$, $P(X_1=r_1, \cdots, X_k=r_k)=\frac{n!}{r_1!\cdots r_k!}p_1^{r_1}\cdots p_k^{r_k}$. 此处 $\sum_{i=1}^{k} r_i=n$ $(\geqslant k)$, $\sum_{i=1}^{k} p_i=1$ $(0<p_i<1)$. 试证:

i) 若损失函数为 $\sum_{i=1}^{k}(p_i-d_i)^2$, 则 $\frac{1}{n+\sqrt{n}}\left(X_1+\frac{\sqrt{n}}{k}, \cdots, X_k+\frac{\sqrt{n}}{k}\right)'$ 是 $(p_1, \cdots p_k)'$ 的容许 minimax 估计.

[提示: 取先验分布密度为 $B^{-1}\left(\frac{\sqrt{n}}{k}; \cdots, \frac{\sqrt{n}}{k}\right)p_1^{\frac{\sqrt{n}}{k}-1}\cdots p_k^{\frac{\sqrt{n}}{k}-1}$.]

ii) 若损失函数为 $\sum_{i=1}^{k}(p_i-d_i)^2\Big/(1-p_i)$, 则 $\frac{1}{n}(X_1, \cdots, X_k)$ 是 $(p_1, \cdots, p_k)'$ 的容许 minimax 估计.

[提示: 取先验分布为 $\sum_{i=1}^{k} p_i=1$ 上均匀分布.]

3. 试求超几何分布

$$P(X=r)=\binom{M}{r}\binom{N-M}{n-r}\Big/\binom{N}{n}$$

在平方损失和 $L\left(d, \frac{M}{N}\right)=\left(d-\frac{M}{N}\right)^2\Big/M(N-M)$ 下的容许 minimax 估计 $\left(\text{规定 } L(0, 0)=L(1, 1)=\frac{N-M}{M(N-1)}\right)$.

[提示: 用先验分布

$$\xi(M=m)=\int_0^1\binom{N}{m}\theta^m(1-\theta)^{N-m}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}d\theta$$

选择适当的 a、b, 使风险函数为常数, 求出 Bayes 估计.]

4. 令 X 有负二项分布 $NB(p, \nu)$, $0<p<1$, $\nu>0$ 已知, 其概率

$$P(X=x)=\frac{\Gamma(x+\nu)}{\Gamma(\nu)\Gamma(x+1)}p^x(1-p)^\nu \quad (x=0, 1, 2, \cdots).$$

i) 试决定 $E_{p,\nu}\{X\}$ 在平方损失下, 哪些 $aX+b$ 是容许的? 在 $0<p_0<p<1$ 限制上给出一容许估计.

ii) 若 p 的先验分布为 Beta 分布 $B(a, b)$, 试求其 Bayes 估计, 此估计是否容许? p 的 MLE 是广义 Bayes 估计吗?是 minimax 估计吗?

5. 令 $X_1, X_2, \cdots, X_n$ 为 n 个 iid. 随机变量, 它有共同密度函数 $f(x, \alpha)=e^{-(x-\alpha)}$ (当 $\alpha\leqslant x<\infty$); $f(x, \alpha)=0$, (其它). 求证 $\hat{\alpha}_n=X_{(1)}-\frac{1}{n}$ 是平方损失下的容许 minimax 估计.

6. 令 X 有 Γ-分布 $g(\theta^{-1}, \nu)$, $\nu>0$ 已知, 求 θ 在平方损失下的容许 minimax 估计, 以及限制 $\theta\geqslant\theta_0>0$ 下的容许估计.

7. 若 $X_1, X_2, \cdots, X_n$ 是 iid 随机变量, 具有极值分布, 密度为 $f(x-\theta)$ $(\theta\in R^1)$ 此处 $f(x)=\exp\{-x-e^{-x}\}$. 试证明 $\sum_{i=1}^n\exp(-x_i)/n$ 是 $e^{-\theta}$ 的容许估计(平方损失). 当损失函数为 $L(e^{-\theta}, d)=(d-e^{-\theta})e^{2\theta}$ 时, 此估计是 minimax 吗?

8. 考虑 logistic 分布族, 密度函数为

$$f(x, \theta)=\exp\{-(x-\theta)\}/\{1+\exp(-(x-\theta))\}^2.$$

i) 决定 θ 的 Pitman 估计, 它在平方损失下是否容许?

ii) 在损失函数 $L(\theta, d)=a(\theta-d)^-+b(\theta-d)^+$ $(a>0, b>0)$ 下的容许 minimax 估计是什么?

9.* 若 k 维随机向量 X 有非退化的有限的协方差阵 $A>0$, 在 $\|d-EX\|^2$ 下, 求 BX 是 EX 在线性估计类中为容许的充要条件(参考[23]).

10.* 若 $X\sim N(\mu, \sigma^2I_k)(k>1)$, $\mu\in R^1$, $\sigma>0$ 皆未知. 试求出 σ^2 在平方损失下, 二次型估计 $X'BX$ $(B>0)$ 在二次型估计类中是容许的充要条件(参考文献[30]).

11.* 试证明在 $X\sim N(\mu, I_k)$、$L(\mu, d)=\sum_{i=1}^k(\mu_i-d_i)^2e^{\mu_i}$ 下, 当 $k\geqslant2$ 时, x 是 μ 的非容许估计(Brown).

[提示: 用恒等式

$$E(x_i-\mu_i)g(x)=E\left(\frac{\partial}{\partial x_i}g(x)\right),$$

$$Ee^{\mu_i}g(x)=E\exp\left(x_i+\frac{1}{2}\right)g(x-e_i),$$

此处 $e_i=(0, \cdots 0, 1, 0, \cdots 0)'$，即在第 i 个坐标为1，其余为零的向量．以此证明 $\hat{\theta}_i=x_i+ce^{-x_i}/\sum_1^k e^{-x_j}$ $(i=1, 2, \cdots, k)$，$0<c<2/e^3$ 优于 x．]

12.* 令 $x_1, x_2, \cdots, x_k$ 为相互独立的随机变量，且 $x_i\sim$ Poisson 分布 $P(\lambda_i)(0<\lambda_i<\infty)$．试证明在损失函数 $\sum_1^n(d_i-\lambda_i)^2/\lambda_i$ 之下 $x=(x_1, \cdots, x_k)'$ $(k\geqslant 2)$ 不是 $\lambda=(\lambda_1, \cdots, \lambda_k)'$ 的容许估计．

[提示：用恒等式 $Eg(x)/\lambda_i=E[g(x+e_i)/(x_i+1)]$ 来证明

$$\hat{\lambda}=\left[1-c/\sum_1^k(x_i+c)\right]x$$

一致优于 x，此处 $0<c<2(k-1)$．]

13．若随机向量 $X_1, X_2, \cdots, X_k$ 彼此独立，X_i 有密度函数 $f(x_i-\theta_i 1)$，且满足定理3.2的条件，记 θ_i 依赖于 x_i 的 Pitman 估计为 $\hat{\theta}_i(x_i)(i=1, 2, \cdots, k)$．则 $\hat{\theta}(x)=(\hat{\theta}_1(x_1), \cdots, \hat{\theta}_k(x_k))'$ 在损失函数

$$\sum_1^k(1+\theta_i^2)^{\frac{r}{2}}(\theta_i-d_i)^2 \quad (r>1)$$

下，它是 $\theta=(\theta_1, \cdots, \theta_k)'$ 的容许估计．问需加何种条件保证在 $r=1$ 时 $\hat{\theta}(x)$ 也是 θ 的容许估计？

[提示：用定理1.5，参考文献[9, ii)]．]

参 考 文 献

[1] J. O. Berger: i) Admissible minimax estimation of a multivariate normal mean with arbitrary quadratic loss, Ann. Statist. 4: 223～236 (1976).

ii) Improving on inadmissible estimators in continuous exponential families with applications to simultaneous estimation of gamma scale parameters, Ann. statist. 8: 345～371 (1980).

[2] C. R. Blyth: On minimax statistical decision procedure and their admissibility, Ann. Math. Statist. 22: 22～42 (1951).

[3] B. Blackwell: On the translation parameter problem for discrete variable, Ann. Math. Statist. 22: 393～399 (1951).

[4] M. E. Bock: Minimax estimators of the mean of multivariate normal distribution, Ann. Statist. 3: 229～258 (1975).

[5] L. Bondesson: When are the mean and the studentized difference independent, Ann. Statist. 4: 668～672 (1976).

[6] I. F. Brewster and J. V. Zidek: Improving on equivariant estimators Ann. statist. 2: 21～38 (1974).

[7] L. D. Brown: i) On the admissibility of invariant estimators of one or more location parameters, Ann. Math. Statist. 37: 1087～1136 (1966).

ii) Admissible estimators, recurrent diffusions and insoluble boundary value problems, Ann. Math. Statist. 42: 855～903 (1971).

iii) Example of Berger's phenomenon in the estimation of independent normal means. Ann. Statist. 8:572～585(1980).

[8] L. D. Brown and M. Fox: Admissibility of procedures in twodimentional location parameter problems, Ann. Statist. 2: 248～266 (1974).

[9] 成平: i) «指数族分布之参数的极小极大化估计», 数学学报 14: 252～275 (1964).

ii) Admissibiliy of simultaneous estimation of several parameters. J. Sys. Sci. & Math. Scis 2(3) (1982), 176～195.

[10] A. Cohen: All admissible linear estimates of the mean vector, Ann. Math. Statist. 37: 458～463 (1966).

[11] M. Clevenson and J. V. Zidek: Simultaneous estimation of the mean of independent Poisson laws, J. Amer. Statist. Asso. 70: 689～705 (1977).

[12] R. H. Farrel: i) Estimation of a location parameter in the absolutely continuous case, Ann. Math. Statist. 35: 949～998 (1964).

ii) On a necessary and sufficient condition for admissibility of estimators when strictly convex loss is used, Ann. Math. Statist. 38: 23～28 (1968).

[13] T. S. Ferguson: Mathematical Statistics: A Decision Theoretic Approach, Academic Press, New York (1967).

[14] M. A. Girshick and L. G. Savage: Bayes and minimax estimates for quadratic loss function, Proc. Second Berkeley Symp. Math. Statist. Prob. I: 53～74 (1951).

[15] L. J. Gleser: Minimax estimation of a normal mean vector when the covariance matrix is unknown, Ann. Statist. 4: 668～672 (1976).

[16] J. L. Hodge and E. L. Lehmann: Some applications of the Cramér-Rao inequality, Proc. Second Berkeley Symp. Math. Statist. Prob. I: 13～22 (1951).

[17] H. M. Hudson: A natural identity for exponential family with applications in multiparameter estimation, Ann. Statist. 6: 473～484 (1978).

[18] W. James and C. Stein: Estimation with quadratic loss, Proc. Fourth Berkeley Symp. Math. statist. Prob. I: 361～379 (1960).

[19] A. M. Kagan: On the estimation theory of locaton parameters, Sankhya Seris A, 28: 335～352 (1966).

[20] A. M. Kagan, Yu. V. Linnik and C. R. Rao: i) Characterization Problems In Mathematical Statistics, John Wiley & Sons (1973).

ii) A characterization of normal law by a admissibility of sample mean,

Sankhya, Seris A 27: 405～406 (1965).

[21] S. Karlin: Admissibility for estimation with quadratic loss, Ann. Math. Statist. 29: 406～436 (1958).

[22] M. W. Katz: Admissible and minimax estimates of parameters in truncated spaces, Ann. Math. Statist. 32: 136～142 (1961).

[23] C. R. Rao: i) Estimation of parameters in a linear model, Ann. Statist. 4: 1023～1037 (1976).

ii) Simultaneous estimation of parameters-a compound decision problem, Statistical Decision Theory and Related Topics II (Edi. by S. S. Gupta and D. S. Moore) 327～ (1976).

[24] Shian-Koong Perng: Inadmissibility of various good statistical procedures which are translation invariant, Tech. Rep. No. 10, Dept. Statist. Michigan State Univ. (1967).

[25] J. Sacks: Generalized Bayes solutions in estimation problems, Ann. Math. Statist. 34: 751～768 (1963).

[26] C. Stein: i) A necessary and sufficient condition for admissibility, Ann. Math. Statist. 26: 518～522 (1955).

ii) Inadmissibility of the usual estimator for the mean of a multivariate normal distribution, Proc. Third Berkeley Sypm. Math. statist. prob. I: 197～206 (1956).

iii) The admissibility of the Pitman's estimator for a single location parameters, Ann. Math. Statist. 30: 970～979 (1959).

iv) Inadmissibility of the usual estimate for the variance of a normal distribution with unknown mean, Annals Inst. Statist. Math. 16: 155～160 (1964).

[27] W. E. Strawdman: Minimax estimation of power of the varianse of a normal population under squared error loss, Ann. Statist. 2: 190～196 (1974).

[28] J. V. Zidek: Sufficient conditions for the admissibility under squared error loss of formal Bayes estimator, Ann. Math. Statist. 41: 446～456 (1970).

[29] A. Wald: Statistical Decision Function, John Wiley, New York (1950) (有中译本).

[30] 吴启光、成平、李国英: «线性模型中误差方差的二次型估计的容许性», 中国科学. 7, 815～825 (1981).

[31] И. П. Натансон: Теория Функций Вещественной Переменной (1957) (有中译本).

名 词 索 引

名 词　　　页 数